

## Research Report

ETS RR–16-26

# Who's on First? Gender Differences in Performance on the SAT® Test on Critical Reading Items With Sports and Science Content

---

Kay Chubbuck

W. Edward Curley

Teresa C. King

August 2016

Discover this journal online at  
**Wiley Online Library**  
wileyonlinelibrary.com

# ETS Research Report Series

---

## EIGNOR EXECUTIVE EDITOR

James Carlson  
*Principal Psychometrician*

## ASSOCIATE EDITORS

Beata Beigman Klebanov  
*Senior Research Scientist*

Heather Buzick  
*Research Scientist*

Brent Bridgeman  
*Distinguished Presidential Appointee*

Keelan Evanini  
*Research Director*

Marna Golub-Smith  
*Principal Psychometrician*

Shelby Haberman  
*Distinguished Presidential Appointee*

Anastassia Loukina  
*Research Scientist*

Donald Powers  
*Managing Principal Research Scientist*

Gautam Puhan  
*Principal Psychometrician*

John Sabatini  
*Managing Principal Research Scientist*

Matthias von Davier  
*Senior Research Director*

Rebecca Zwick  
*Distinguished Presidential Appointee*

## PRODUCTION EDITORS

Kim Fryer  
*Manager, Editing Services*

Ayleen Gontz  
*Senior Editor*

---

Since its 1947 founding, ETS has conducted and disseminated scientific research to support its products and services, and to advance the measurement and education fields. In keeping with these goals, ETS is committed to making its research freely available to the professional community and to the general public. Published accounts of ETS research, including papers in the ETS Research Report series, undergo a formal peer-review process by ETS staff to ensure that they meet established scientific and professional standards. All such ETS-conducted peer reviews are in addition to any reviews that outside organizations may provide as part of their own publication processes. Peer review notwithstanding, the positions expressed in the ETS Research Report series and other published accounts of ETS research are those of the authors and not necessarily those of the Officers and Trustees of Educational Testing Service.

The Daniel Eignor Editorship is named in honor of Dr. Daniel R. Eignor, who from 2001 until 2011 served the Research and Development division as Editor for the ETS Research Report series. The Eignor Editorship has been created to recognize the pivotal leadership role that Dr. Eignor played in the research publication process at ETS.

## RESEARCH REPORT

# Who's on First? Gender Differences in Performance on the SAT® Test on Critical Reading Items With Sports and Science Content

Kay Chubbuck, W. Edward Curley, & Teresa C. King

Educational Testing Service, Princeton, NJ

This study gathered quantitative and qualitative evidence concerning gender differences in performance by using critical reading material on the SAT® test with sports and science content. The fundamental research questions guiding the study were: If sports and science are to be included in a skills test, what kinds of material are appropriate for fair and valid assessments? Does item type matter with regard to differential item functioning (DIF)? What factors can provide explanations for the different levels of difficulty, discrimination, and DIF exhibited? Do differences in students' interest in and familiarity with the material affect their performance? From data gathered during the two phases of our research, it appears that most sports and science material of the sort used in this study should be permissible in a skills test. Particularly in the context of a long passage, in which most information needed to answer an item is provided, very little material appears to be too "specialized." From participants' statements during cognitive interviews, it seems that interest in and/or familiarity with a subject has little impact on performance—despite previous research to the contrary. On the basis of our observations, we recommend the following: (a) use more than one type of statistic to evaluate differential item functioning; (b) consider the amount of context provided with an item rather than just the particular content of an item; and (c) when possible, use data rather than assumptions about gender bias to make decisions about test content that may generate differential item functioning.

**Keywords** SAT®; DIF; STD-EISDIF; Mantel–Haenszel; gender; reading; sports; science

doi:10.1002/ets2.12109

The SAT® test is taken primarily by college-bound juniors and seniors in U.S. high schools. Rather than measuring knowledge of particular content (such as biology or history), the SAT is a skills test whose critical reading sections (SAT-CR) have historically measured knowledge of the meanings of words; the ability to understand how sentences fit together logically; and the ability to understand, analyze, evaluate, and draw inferences from reading passages of varying lengths. At the time of this research, the SAT-CR included two item types (sentence completions and passage-based reading questions), and the material in the SAT-CR was not determined by any specific curriculum. Instead, the words, sentences, and reading passages were intended to be fair and valid for high school juniors and seniors, related to their future success in college.

A skills test that is fair and valid for predicting performance in college for its intended population as a whole should also be fair and valid for the gender and ethnic groups of examinees who comprise that population. In addition to traditional item analysis (IA), which measures the difficulty and discrimination of test questions for the entire sample, differential item functioning (DIF) statistics can be used to identify test questions on which groups of examinees, matched for ability (as indexed by total test scores), perform differently. During the 1980s and 1990s, many studies of the SAT were conducted at Educational Testing Service (ETS) and elsewhere to analyze and report on sources of DIF for various groups of examinees (see Bleistein, Schmitt, & Curley, 1990; Bleistein & Wright, 1987; Schmitt, 1988; Schmitt, Curley, Bleistein, & Dorans, 1988; Schmitt & Dorans, 1990). This research informed both the procedures used at ETS to evaluate DIF and the practices documented in the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2009). This same research has served to direct item writing and test assembly for the SAT-CR over the past 35 years.

Specifically, the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2009) urged test developers to “avoid cognitive sources of construct-irrelevant variance.” To this end, Guideline 1 states that skills tests should “avoid unnecessarily specialized vocabulary unless such vocabulary is important to the construct being assessed.” The very next sentence, however, makes an important point: “What is considered unnecessarily specialized requires judgment.” The same holds true for the

*Corresponding author:* K. Chubbuck, E-mail: kchubbuck@ets.org

various topics that are covered in the reading passages that appear on skills tests such as the SAT-CR. They should not be unnecessarily specialized; however, professional judgment is required to make such distinctions, and whenever possible, such judgments about vocabulary and reading passages should be informed by statistical data and the occasional qualitative information that are gathered directly from the population of examinees who actually take the examination.

The question of what constitutes *specialized* is, therefore, paramount for a skills test such as the SAT-CR. The purpose of this study is to explore this question by analyzing material in two content areas of interest: sports and science. Since 1986, sports content has been largely avoided in the SAT, in part because of historically different access by gender to school sports and a presumption that females would therefore be disadvantaged during testing if such material were to be included. Willingham and Cole (1997) used data gathered in the 1980s to stipulate that sports content should be avoided in skills testing: “[F]ootball in the case of women” is “irrelevant to what is being measured but may affect performance because that context is less familiar” (p. 363). Differences in sports interests and experiences had led to differing performance on items with a sports context; therefore, prohibiting sports content in the SAT seemed a reasonable decision at that time in the interest of promoting test fairness.

By 2007, however, Title IX had brought about major changes. Female participation in high school athletics had increased 904% and female participation in college athletics had increased 456% since the federal law went into effect in 1975, per the Women’s Sports Foundation (2008), suggesting a transformation so thorough as to call into question the continued relevance of earlier observations. This change is reflected in the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2009), which stated that sports content may now be used in skills testing provided that the item does not require specialized knowledge, such as how many points a field goal is worth. Nonetheless, some test developers remain skeptical about the inclusion of sports content, even in general terms. As recently as 2013, sports were still being singled out in mandatory fairness training sessions for test developers at ETS as a possible source of construct-irrelevant variance out of concern that such material might discriminate against female test takers.

Unlike sports content, science content has long been a part of the specifications for SAT-CR. As Nancy Cole noted in *The ETS Gender Study* in 1997, the gender difference in performance on science items has for some time been “essentially zero. The familiar math and science advantage for males ... [is] significantly smaller than 30 years ago” (p. 3). Indeed, both male and female test takers can be presumed to have significantly greater familiarity with science than in the past. According to research conducted by the College Board (CB), 61% of SAT test takers in 2011 reported 4 or more years of science course work, compared with 40% of SAT test takers in 1991. Similar changes were not observed in other subjects. For example, there was a 2% decrease in the number of students reporting 4 or more years of English/language arts course work during the same period (SAT Update, 2011). Yet in selecting science passages for skills testing, consideration has long been given to whether such passages might be deemed unduly specialized. Additionally, a pattern of DIF against females on sentence completion (SNCP) items with science content on the SAT-CR has persisted. In 2010, a 5-year analysis of SAT-CR SNCPs that had yielded extreme levels of DIF showed that 57% of items that had disadvantaged females included science content in either the stem or the options, with a particular pattern of difference surrounding the words *camouflage*, *eclipse*, *migratory*, *plateau*, *pioneer*, and *velocity* when those words were used as the item keys. Such data call into question why female test takers performed differently than male test takers when matched for ability. Was some aspect of the item type causing the variance, as opposed to the item content alone? Or was it simply a lack of familiarity with certain terms or topics from the domain of science, about which it would be hard to generalize?

To address these concerns, this study was designed to gather both quantitative and qualitative evidence concerning real and perceived differences in performance between males and females on SAT-CR material with sports and science content. The following are the fundamental research questions guiding the study:

- If sports and science are to be included in the SAT-CR, what kinds of material are appropriate on which to base fair and valid score-based inferences?
- What constitutes specialized material for these content areas?
- Does item type matter with regard to DIF? Namely, are there observable differences in performance on SNCP versus reading-based items as well as on reading-based items for short passages versus long passages?

### Phase 1: Quantitative Study, 2009–2010

With CB approval and ETS Research and Development funding support, a range of SAT-CR sports and science material was developed for testing in 2009 and 2010. Passage-based reading questions and a variety of sports and science SNCPs



**Table 1** SAT-CR Material Approved and Developed for the Study

	Sports	Science
Sentence completions	17	20
Short reading passages	2	4
Long reading passages	3	1
Total passage-based items	57	32

were written, reviewed, and approved. This included all required content reviews and fairness approvals. Before any material can appear in a test, its content must be approved as being in alignment with the *ETS Guidelines for Fairness Review of Assessment* (ETS, 2009). The sports and science material in each item type that was part of this study during its 2-year development cycle is shown in Table 1. Note that some of the item types have relatively small counts; this was a limitation of the study.

In addition, a single *variant* of some of the sports and science SNCPs was developed during 2009, the first year of the study. These items maintained the sentence structure and idiom of the original item and included identical answer choices, but the content was changed from sports or science to a more neutral subject for the purposes of identifying, as much as possible, whether it was the content of the item or the vocabulary in the answer choices that most affected the item's statistics. Samples are provided in the "Results and Discussion" sections for this phase of the report.

All of the study items were embedded in nonscored variable sections of the SAT, along with other nonstudy new material being tried out to collect statistical data. Each of these sections was taken by approximately 3,500 test takers representing the typical SAT population in terms of gender and ethnicity. (It was not part of this study to examine crossover DIF effects, although those data were gathered on the items.) Great care was taken to ensure that none of these nonscored sections had an inappropriate amount of sports or science content. For those SNCPs that had variants, the variants were located in parallel nonscored sections and in identical item positions as the original SNCPs on which they were based. Each set of passage-based reading items was likewise split into two groups of items and administered in parallel nonscored sections along with the study passages, just as nonstudy new reading passages and items are tried out to collect statistical data on their performance prior to being used in operational testing. In this way, every nonscored section looked like an operational section in an effort to ensure appropriate examinee motivation as well as comparable difficulty and speededness of sections.

Just as each passage and item received all of the usual reviews and approvals, each of the nonscored sections in which study items appeared had all of the usual ETS and CB content reviews before being tried out as part of an operational exam package. Each section was administered to a spaced random sample of at least 3,500 SAT examinees along with the full regular operational form of the test. After the administration, data on the study items were analyzed in three different ways:

- traditional IA to determine item difficulty and discrimination;
- traditional Mantel–Haenszel (MH) DIF analysis (Holland & Thayer, 1988) to determine MH values and ETS DIF categories for each item; and
- standardized (STAND) DIF analysis (Dorans & Kulick, 1983, 1986) to provide additional information about the performance of males and females (as well as ethnic group members) on the study items.

Both the MH and STAND methods identify DIF after partitioning the reference group (in this study, males) and the focal group (in this study, females) into subgroups with the same score on a relevant matching variable (in this study, the regular operational SAT-CR form). While there are some differences between the MH and STAND methods (Dorans, 2013; Dorans & Holland, 1992), the DIF estimates computed by the two methods are highly correlated (in the upper .90s); they tend to yield the same rank order of items with respect to DIF (Dorans, 1989; Holland & Thayer, 1988; Wright, 1987). Moreover, by including only those examinees who reached an item in the calculation of the DIF value for that item, an attempt is made to mitigate the effects of differential speededness on the values of the DIF indices (Schmitt & Bleistein, 1987).

In the STAND analysis, an item is said to exhibit DIF when the probability of correctly answering the item is lower or higher for examinees from one group than for equally able examinees from another group when matched for ability as judged by performance on the test as a whole. The basic elements of a STAND analysis are proportions correct for

the reference and the focal groups at each level of the matching criterion. The standardized difference in expected item performance (STD-EISDIF) index quantifies DIF. It can be expressed in the  $p$  metric (i.e., percentage of examinees who selected an answer choice). In this metric, the index can range from  $-1$  to  $+1$ , or from  $-100\%$  to  $100\%$ . Negative values of STD-EISDIF indicate that the item favors the reference group, whereas positive values indicate that the item favors the focal group. STD-EISDIF values between  $-0.05$  ( $-5\%$ ) and  $+0.05$  ( $5\%$ ) are considered negligible. STD-EISDIF values outside the  $-0.10$  and  $+0.10$  ( $-10\%$  and  $10\%$ ) range merit careful investigation. For exploratory research purposes, a cutoff of  $|\text{STD-EISDIF}| \geq 0.05$  is sometimes used to identify items for investigation. In addition to calculating STD-EISDIF values for the *key* (correct answer choice) to an item, differences in the standardized proportion of responses for each *distractor* (incorrect answer choice) are also computed and studied to clarify the effects of the hypothesized DIF factors (Dorans, Schmitt, & Blestein, 1992). The STD-EISDIF uses all available males and females to “match” males to the female group.

The MH procedure for DIF analysis computes ratios of the conditional odds of successful reference group performance over the conditional odds of successful focal group performance at each score level and then averages these ratios across all score levels. In the calculation of the average ratio, statistically optimal weights are used for each ratio. The MH method provides an estimate of the constant odds ratio. The MH statistic is transformed to the delta metric, which is used to indicate item difficulty in the ETS test development process. To obtain a delta, the proportion correct ( $p$ ) is converted to a  $z$ -score via a  $p$ -to- $z$  transformation using the inverse of the normal cumulative function, followed by a linear transformation to a metric with a mean of 13 and a standard deviation of 4. Large values in a delta metric, such as 16 or 18, correspond to difficult items, whereas easy questions have small delta values, such as 7 or 9. This is the opposite of proportion correct in that higher does not equal easier. The MH estimate of DIF effect size in the delta difference metric (MH D-DIF) ranges from negative infinity to infinity, with a value of 0 indicating no DIF. MH D-DIF values between  $-1.00$  and  $+1.00$  are considered negligible. MH D-DIF values outside the  $-1.50$  to  $+1.50$  range merit careful investigation. For exploratory research purposes, a cutoff of  $|\text{MH D-DIF}| \geq 1.00$  may be used. As with the STD-EISDIF index, positive values of MH D-DIF favor the focal group, whereas negative values disadvantage the focal group. Note that MH D-DIF can be calculated only for the key of each item, not for the distractors, which limits the usefulness of the MH statistic for research such as that being conducted in this study. For a complete description and comparison of the STD-EISDIF and MH D-DIF statistics, refer to Dorans and Holland (1992).

In the present study, DIF categorization of items was made on the basis of the standard ETS DIF operational item screening classifications (Petersen, 1988). These classifications are as follows:

- A items have an MH D-DIF not significantly different from zero (at the .05 level) or an absolute value less than 1.00.
- C items have an absolute value of MH D-DIF of at least 1.50 and significantly greater than 1.00 (at the .05 level).
- B items are those that are neither A nor C.

Note that the analysis of DIF for the SAT involves a two-step process to refine the matching criteria. First, the total test raw score on the SAT-CR operational form is used as the matching criterion to determine DIF for each of the 67 operational items that appear on that test form. Second, on the basis of that initial analysis, any operational item with an extreme DIF value (C DIF) favoring the reference group in a comparison with a particular focal group is removed as part of the total score that is used to match that particular focal group with the reference group. Thus a refined matching criterion is determined for each focal group comparison for use in the subsequent DIF analysis.

## Results and Discussion

### *Sports Content in Sentence Completion Items*

Seventeen SNCPs with sports content were administered as part of this study, and only one of those items showed a large (16%) difference in performance between matched groups of females and males. Five other SNCPs with sports content were classified as C DIF against females using the MH statistic, but four of those five were easy items for the total population (ranging from  $\delta = 6.9$  to  $8.3$ ), and the fifth was very hard ( $\delta = 16.5$ ). All five of these easy and hard items that were classified as C DIF actually showed small (5–6%) differences between matched groups of females and males when using the STD-EISDIF statistic. It should be noted that categorizing such items as C DIF occurs because the MH D-DIF is unbounded at the two ends of the scale, unlike the STD-EISDIF statistic, which represents differences in proportions correct on a bounded scale of 0–100 (Dorans & Holland, 1992). Although the DIF estimates derived by these two methods are highly

correlated, they tend to differ most often with the easiest and the hardest SAT items. MH D-DIF on these kinds of items is sometimes called *artifactual* because the C DIF classification does not reflect a sizable difference in the performance of the focal group and the reference group. Rather, the classification is an artifact (i.e., an unintended or misleading effect) of the unbounded scale, representing a difference of less than 10%.

The remaining 11 SNCPs with sports content showed A DIF or B DIF for the male–female comparison and covered the full range of difficulty ( $\delta = 5$  to 15). All but one of the 17 sports SNCPs had acceptable discrimination for the total population. The  $\delta = 15$  item had an  $r$ -biserial of .09, which indicates that the item did not discriminate appropriately among the higher and lower scoring students based on the test takers' performance on the operational sections of the SAT-CR. The  $r$ -biserial is the correlation between performance on an individual item and performance on the overall set of items (the exam). An item with an  $r$ -biserial lower than .2 is not discriminating in helpful ways among the more able and less able test takers.

Table 2 includes 11 of the 17 SNCPs with sports content: first, the item with the greatest male–female C DIF, then the five items with artifactual male–female C DIF (i.e., small differences in performance), and finally, five of the remaining 11 items that showed A DIF or B DIF for the male–female comparison.

It should be noted, first of all, that none of these research items is overly specialized in its discussion of sports; none includes specialized sports terminology in the answer choices (although there is some sports terminology in the question stems); and none requires detailed knowledge of the rules, strategies, or history of a particular sport. As stated previously, the SAT-CR is a general skills test that should not measure specialized content knowledge in any of its items. The question is whether the mere reference to a sport, an athlete, or a particular sports-related issue will adversely affect the performance of females matched in ability with males. Our analysis of these 17 SNCPs seems to provide evidence that the mere presence of sports terminology did not result in DIF at a level that would cause concern.

Where the sports context did seem to matter was in the stem of one of the items in which sports vocabulary was used to establish the context for the answer options. This usage led to the greatest difference between males and females in performance when matched for ability. More specifically, in the first item in Table 2, the statement, “did not hesitate to body check opponents into the boards,” points directly to the key, a “physical” style of play—but to understand this connection, test takers needed to be able to decode the phrase “body check” as meaning more than just careful observation. This task seems to have been more challenging for female than for male test takers: There was a 16% difference. The data indicate that matched females who did not select the key (E) primarily chose option (A) “methodical” and option (C) “cunning”—6% more often than matched males for each of these options. This suggests that “body check” is a term with which female test takers are less familiar than male test takers are, or it is a term that they interpret differently than males when matched for ability. See the section, “Phase 2: Qualitative Study, 2011–2012,” for a more detailed discussion of this item.

That said, the use of sports terminology in the stems of items did not always produce such stark differences in performance. For example, in the second item in Table 2, the statement “play 18 holes of golf in ten under par one day and fail to break par the next day” could also be construed as moderately technical, yet matched females selected the key, “inconsistent,” only 5% less often than matched males did. The item is classified as C DIF, but the DIF is artifactual. Nor did females select any of the incorrect choices significantly more often than matched males did. Notably, this discussion about golf is longer than the phrase “body check,” and the inclusion of the word “fail” with regard to the golfer's performance on the second day as compared to the golfer's performance on the first day may have helped test takers to select the key “inconsistent,” regardless of their familiarity with the topic.

Even more notable are the results for the football item, the ninth item in Table 2. Although football is not mentioned explicitly in the item, the stem is full of specific terms and descriptions about football: “star wide receiver,” “over 80 catches,” “passing skills,” and “an incredibly strong and accurate throwing arm.” Yet the item was A DIF, showing only a 3% difference in performance between matched females and males. Again, it is possible that the increased context provided in the stem helped test takers to identify the key regardless of the sports context.

A similar statistical pattern is seen in several other items in Table 2. In the field hockey item, the key (C) draws only 5% fewer females than males when matched for ability while distractor (D) draws only 3% more females than males. In the short-track skating item—a very difficult question answered correctly by only 13% of the total population—the key (E) draws only 6% fewer females than males when matched for ability, whereas 3% more females than males chose to omit the item. Similar results are seen with the bobsledding and marathon items.

**Table 2** Eleven SAT-CR Sentence Completion Items With Sports Content Administered in the Study

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
G8-06	2	8.9	0.47	-2.96	C		Known for his — — style of play, James did not hesitate to body check opponents into the boards as he shot the puck across the ice toward the hockey goal. (A) methodical (B) inconsistent (C) cunning (D) adaptable (E) physical <sup>b</sup> (OMITS)
G8-30	2	7.2	0.64	-1.68	C		Performance on the golf course is often — — : a professional golfer can play 18 holes in ten under par one day and fail to break par the next day. (A) enjoyable (B) profitable (C) inconsistent <sup>b</sup> (D) uninteresting (E) deliberate (OMITS)
F5-30	2	8.3	0.69	-1.51	C		Although the precise — — of field hockey remains a mystery, 4,000 year-old drawings found in the Nile River valley of Egypt — — men apparently swinging sticks at a round object. (A) contribution ... forecast (B) purpose ... denounce (C) origin ... depict <sup>b</sup> (D) significance ... portray (E) fate ... illustrate (OMITS)
G8-24	4	16.5	0.51	-1.76	C		Injuries are — — among short-track skaters: they occur so routinely that short track is often described as roller derby on ice. (A) nondescript (B) mercurial (C) tractable (D) puerile (E) rifle <sup>b</sup> (OMITS)

Table 2 Continued

Test	Item number	Equated delta	r-biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-18	4	6.9	0.41	-2.61	C		The bobsled team reached a peak speed of 120 miles per hour, a — that almost set a new world record. (A) vulnerability (B) longevity (C) velocity <sup>b</sup> (D) durability (E) dependability (OMITS)
G8-04	4	7.0	0.55	-2.32	C		With an official distance of more than 42 kilometers, the marathon is a true test of —: it takes most runners many grueling hours to complete. (A) dexterity (B) teamwork (C) opportunity (D) stamina <sup>b</sup> (E) independence (OMITS)
F5-32	5	10.9	0.66	-1.04	B		Though China's Shaolin temple is celebrated primarily as the birthplace of kung fu, most other martial arts can — their — to the ancient monastery as well. (A) expose ... ancestry (B) compare ... inception (C) attribute ... dissolution (D) demonstrate ... superiority (E) trace ... genesis <sup>b</sup> (OMITS)
F5-26	7	14.5	0.35	1.10	B		While competing in the Tour de France, the cyclist was remarkably —: on the bike, he was unsmiling, and during public appearances, he was — to the point of being unresponsive. (A) accessible ... cagey (B) aloof ... reticent <sup>b</sup> (C) supercilious ... deferential (D) gregarious ... taciturn (E) cantankerous ... voluble (OMITS)

Table 2 Continued

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-24	3	7.9	0.48	-0.97	A		As a star wide receiver with over 80 catches this year, Smith rarely has a chance to — his passing skills in games, but he has an incredibly strong and accurate throwing arm.
							(A) placate
							(B) exhibit <sup>b</sup>
							(C) begrudge
							(D) restrict
							(E) neglect
F5-20	2	11.3	0.56	-0.54	A		(OMITS)
							Already a widely recognized figure in the field of gymnastics, Nadia Comaneci attained worldwide — when she became the first gymnast to score a perfect ten in Olympic competition.
							(A) disgrace
							(B) prudence
							(C) renown <sup>b</sup>
							(D) integrity
F5-38	7	10.4	0.42	-0.03	A		(E) humility
							(OMITS)
							Baseball player "Cool Papa" Bell — that he was so fast, he could turn out the lights and be in bed before the room got dark, a claim that was clearly —.
							(A) confessed ... anecdotal
							(B) bragged ... circumspect
							(C) boasted ... hyperbolic <sup>b</sup>
							(D) quibbled ... spurious
							(E) professed ... credible
							(OMITS)

*Note.* DIF = differential item functioning; MH D-DIF = Mantel-Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup> Values do not always sum to zero because of rounding. <sup>b</sup> Key, or answer, to each item.



It is worth noting that some sports items appear to have advantaged female test takers, though this may have had little to do with the sports content provided in the item. For example, the SNCP in Table 2 about the cyclist in the Tour de France is of above average difficulty ( $\delta = 14.5$ ) for the total population, with 9% more females than matched males selecting the key (B). Yet how much does answering this item correctly rely on the specific context of the Tour de France? In this case, it seems to have been more critical to understand the description of the demeanor of the cyclist and to be familiar with the words “aloof” and “reticent”—factors that seem to have advantaged females over males when matched for ability. For this item, the sports context seems less relevant than the ability to decipher human behavior—an arena in which females have traditionally outperformed males (Halpern, 1992). Something similar could perhaps be said for the sports item about “Cool Papa” Bell. It shows no difference in performance between males and females when they are matched for ability, yet this item too is more about human behavior than about athletics.

As mentioned previously, some SNCPs had variants that were part of this study. Table 3 shows two pairs of these variants, which are representative of the others for which data were collected. The first and third items in Table 3 also appear in Table 2. They are reproduced in Table 3 for ease of comparison with their variants.

As can be seen in the first pair of variants, the reference to “bobsled team” was changed to “new aircraft,” and the reference to “new world record” was changed to “a major selling point.” Otherwise, these first two SNCPs in Table 3 are structurally similar, and the five answer choices are identical in each. The STD-EISDIF data differ hardly at all: only 5% versus 3% fewer matched females than males chose the correct answer. Similarly, for the other pair of variants, the “star wide receiver” becomes “a professional violinist,” “passing skills” becomes “singing skills,” and the “strong and accurate throwing arm” becomes a “strong and beautiful voice.” Otherwise, these pairs are structurally similar, and the answer choices are identical. As with the first pair of variants in Table 3, the STD-EISDIF values are almost indistinguishable: 3% versus 2% fewer matched females than males answered correctly. Even though there is artifactual C DIF in the first item, it seems from the STD-EISDIF values that the sports context in the stems of the two items results in very little actual difference in performance between matched groups of females and males.

### **Sports Content in SAT Critical Reading Passages**

Five SAT reading passages with sports content and 57 related passage-based items were administered as part of this study, as shown in Table 6. The passages were selected to span a range of content, not just in terms of the type of sport referenced but also in terms of the level of sports-specific detail that was included. In this context, *nonspecialized* means the passage includes little to no detail about how the sport is played; *moderately specialized* means it includes some detail; and *more than moderately specialized* means the passage includes much detail and/or vocabulary about the particular sport in the content of the passage.

Of the 57 passage-based items related to the five passages with sports content, only two items show C DIF against females. Four other items show B DIF against females, but only one of the four shows a difference in performance greater than 10%, and that one is just 11% against females matched for ability. The remaining 51 sports-related passage-based reading items show A DIF or relatively minor positive or negative B DIF for the male–female comparison. The 57 reading items include a full range of difficulty ( $\delta = 7$  to 16), and the number of items with acceptable discrimination for the total population is 48 (84% of the total 57), which is very similar to the percentage that is usually derived for pretest (nonstudy) SAT reading sets.

Both C DIF items are associated with the short baseball passage about Wee Willie Keeler, and the B DIF item that was answered correctly by 11% fewer females than matched males is associated with the long women’s boxing passage. Table 4 includes all four of the items associated with Wee Willie Keeler; Table 5 includes a representative sample of the 16 items associated with the long women’s boxing passage, including the one B DIF item.

The short baseball passage about Wee Willie Keeler is *more than moderately specialized* in that it refers to specific aspects of baseball (e.g., “an astonishing .432 batting average”) and to how baseball evolved over the years (e.g., “Every pitch is charted, every hit mapped to the nearest square inch”). The first item in Table 6 does not focus on such details, however, asking instead for examinees to identify the overall purpose of the passage. Females selected the correct answer, “describe a significant transformation in a sport,” 7% less often than males did, when matched for ability. This is not a large difference, but in part, because the item is easy for the overall population, it is classified as C DIF against females. Females chose option (B) 3% more often than matched males did. Option (B) mentions “changes,” but it is incorrect because of the word “condemn.”



**Table 3** Two SAT-CR Sentence Completion Items With Sports Content and Variants Administered in Study

Test	Item number	Equated delta	<i>r</i> -biserial	MHD-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-18	4	6.9	0.41	-2.61	C		The bobsled team reached a peak speed of 120 miles per hour, a — that almost set a new world record. (A) vulnerability (B) longevity (C) velocity <sup>b</sup> (D) durability (E) dependability (OMITS)
F5-20	4	9.4	0.38	-0.49	A		The new aircraft has reached a peak speed of 600 miles per hour, a — that is a major selling point. (A) vulnerability (B) longevity (C) velocity <sup>b</sup> (D) durability (E) dependability (OMITS)
F5-24	3	7.9	0.48	-0.97	A		As a star wide receiver with over 80 catches this year, Smith rarely has a chance to — his passing skills in games, but he has an incredibly strong and accurate throwing arm. (A) placate (B) exhibit <sup>b</sup> (C) begrudge (D) restrict (E) neglect (OMITS)
F5-26	3	7.2	0.51	-1.10	B		As a professional violinist in the symphony orchestra, Smith rarely has a chance to — her singing skills in public, but she has an incredibly strong and beautiful voice. (A) placate (B) exhibit <sup>b</sup> (C) begrudge (D) restrict (E) neglect (OMITS)

*Note.* DIF = differential item functioning; MH D-DIF = Mantel-Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

**Table 4** Critical Reading Items on Wee Willie Keeler Short Passage

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-18	9	8.6	0.45	-1.56	C	1 3 -7 1 1 1	The passage serves primarily to (A) express nostalgia for a bygone time (B) condemn changes in the way a game is played (C) describe a significant transformation in a sport <sup>b</sup> (D) advocate the continuation of a traditional practice (E) celebrate the achievements of two talented athletes (OMITS)
							According to the passage, Boggs's opponents were (A) less experienced than modern baseball players (B) less consistent than earlier baseball players (C) more talented than their predecessors <sup>b</sup> (D) better prepared than Keeler's opponents <sup>b</sup> (E) better compensated than Keeler's contemporaries (OMITS)
							The function of the parenthetical phrase in lines 3–4 is to (A) defend the accuracy of a statistical claim (B) express skepticism concerning a long-standing record (C) indicate an exception to an earlier generalization (D) highlight an interesting but irrelevant point <sup>b</sup> (E) provide evidence in support of a general assertion <sup>b</sup> (OMITS)
							It can be inferred from the passage that the author believes that Keeler (A) succeeded because his opponents lacked the physical talents of contemporary athletes (B) was more inherently athletic than modern baseball players (C) used scientific insights to gain an advantage over other players (D) would not have had as good a batting average if he had played in Boggs's era <sup>b</sup> (E) might have performed even better had he played today (OMITS)
F5-18	10	10.6	0.50	-1.14	B	4 3 -3 -9 1 3	
F5-20	9	11.4	0.42	-0.39	A	-1 -1 1 3 -3 1	
F5-20	10	12.0	0.48	-1.61	C	2 4 3 -14 4 1	

*Note.* DIF = differential item functioning; MH D-DIF = Mantel–Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

Table 5 Critical Reading Items on Women's Boxing Long Passage

Test	Item number	Equated delta	r-biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
G8-30	16	12.5	0.35	0.44	A	4	The "physical reminder" (line 4) refers to the author's (A) sensory experience of boxing <sup>b</sup> (B) nostalgic memories of boxing (C) old injuries from boxing (D) boxing trophies and souvenirs (E) boxing garb and gear (OMITS)
G8-30	17	7.8	0.54	0.81	A	0	The discussion in lines 5–11 suggests that (A) the author boxes every day (B) the author is uncertain why she boxes (C) the author finds boxing draining (D) boxing is central to the author's sense of self <sup>b</sup> (E) boxing is an especially arduous sport (OMITS)
G8-30	19	11.4	0.42	–1.28	B	–1	In lines 19–23 ("He said ... fight"), John suggests that the two boxers are both (A) stylish (B) capricious (C) antagonistic (D) tenacious <sup>b</sup> (E) petulant (OMITS)
G8-30	20	9.5	0.48	–0.76	A	5	Lines 24–61 are focused primarily on (A) a boxing lesson (B) a boxing match (C) a training workout <sup>b</sup> (D) an exercise class (E) an exciting dream (OMITS)
G8-30	21	12.1	0.59	–0.69	A	0	The "story" mentioned twice in line 32 refers most directly to the author's (A) manuscript about the sport of boxing (B) fantasy about being a professional boxer (C) pretense that she is skilled at boxing (D) match with an invented adversary <sup>b</sup> (E) internal monologue to help reinforce her self-esteem (OMITS)
G8-30	22	14.3	0.55	0.09	A	–2	The description in lines 44–46 ("The spun ... smoke") serves to convey (A) the drama of an initial conflict (B) the perspective of an outside observer (C) the poignancy of a chance encounter (D) the symbolism of key images (E) the enjoyment of a momentary respite <sup>b</sup> (OMITS)

Note. DIF = differential item functioning; MH D-DIF = Mantel–Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance. Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

**Table 6** SAT-CR Passages With Sports Content Administered in Study

Passage	Word length	Number of items
Pond Hockey (nonspecialized)	109	4
College Football (nonspecialized)	486	12
Women's Boxing (moderately specialized) <sup>a</sup>	693	16
Wee Willie Keeler (more than moderately specialized) <sup>a</sup>	108	4
Umpires (more than moderately specialized)	841	21

<sup>a</sup>Reproduced in Appendixes A and B. These are the two sports reading passages for which some or all of the related items and their data are included in this report and in the cognitive interviews conducted in 2011.

The second item on the short baseball passage is B DIF, but it shows 9% fewer females than matched males choosing the key (D). This item, like the fourth item in Table 4, requires a deeper understanding of the specific points made in the passage than do the first and third items about the primary purpose of the passage and the function of the parenthetical phrase in Lines 3–4. Thus, given the *more than moderately specialized* nature of the passage, it is perhaps not surprising that the second item and the fourth item in Table 4 show larger differences in the performance of matched females and males than do the first and the third items. None of the incorrect choices in these items attracted a much larger proportion of females than males, and females did not omit these two items proportionally more often; rather, they spread themselves fairly evenly across most of the incorrect choices. From these data, it seems that some females did not understand the fundamental comparison of the two baseball eras as well as males did, when matched for ability.

Notably, such a pattern of gender difference on items related to passages with sports content is found only in this short baseball passage. Three of the six sports reading items that showed C DIF or B DIF against females appear in this set. Said another way, across the 57 items associated with the five research passages with sports content, half of the items that showed any appreciable DIF against females are associated with this short baseball passage, suggesting either that there is something unusual about this passage or that the combination of sports content with short length may have been responsible for the difference. For example, in comparing the Wee Willie Keeler passage (108 words) to the similarly specialized yet significantly longer passage on umpires (841 words), it is notable that none of the items associated with umpires yielded C DIF against females. Of the 21 items associated with umpires, 19 were A DIF for females and 2 were B DIF, with one of the B DIF items registering as advantaging females over males when matched for ability. That particular item characterized the presentation of baseball in the passage as “engrossing.” The passage on umpires, like that about Wee Willie Keeler, contains a discussion of the sport that is *more than moderately specialized*, using language such as “arm signals for balls and strikes,” “the Sultan of Swat,” and “the Babe Ruth of umpiring” as well as a scenario in which fans blame an “erroneous sixth-game, ninth-inning call at first base for robbing the Cardinals of the 1985 World Series title” and a colorful declaration, “It ain’t nothing until I call it!” Having such language dispersed across 841 words, however, seems to have provided more opportunity for decoding the overall meaning of the passage than was available in the more compressed 108 words of the Wee Willie Keeler passage about a similar topic—unless the difference arose because of other factors, such as the inclusion of numerals and decimal points in the Wee Willie Keeler passage to describe batting averages. See the section, “Phase 2: Qualitative Study, 2011–2012,” for additional discussion of this specific aspect of the passage.

Table 5 includes a representative sample of the items based on the long women’s boxing passage. This passage is *moderately specialized* in that it describes in some detail a boxing workout, including references to shadowboxing, footwork, the heavy bag, and the speed bag. It also mentions some actual sparring between the narrator and her friend. Yet the language used to describe boxing in the passage is not as specialized as the language used to describe baseball in the Wee Willie Keeler or the umpires passages. With one exception, the 16 items associated with this passage show little to no difference in performance between matched females and males. The third item in Table 5, which is focused on Lines 19–23, shows that 11% fewer females than matched males selected the correct answer, “tenacious.” Instead, females selected incorrect options (B), (C), and (E) more often than matched males did, and 5% more females than matched males chose to omit the question. There is no difference less than 3% in magnitude for this item.

With regard to the rest of the items in the women’s boxing set, females performed better than males on some and males performed better than females on others when matched for ability, but the differences are not more than 5–6%. The final item in Table 5 shows no difference in performance at all. Overall, the subject matter of this passage did not seem to advantage or disadvantage females. Gender differences in performance were negligible.

### Science Content in Sentence Completion Items

Twenty SNCPs with content and/or vocabulary drawn from a variety of science domains were part of this study. Keys ranged from common words listed in the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2009), such as “microscopic,” to vocabulary that could be considered at least *moderately specialized*, such as “seismic,” “barometer,” and “camouflage,” as a means of examining a range of scientific complexity. Of these 20 items, some included science content in the stem and science vocabulary in the options, some had science content in the stem but no science vocabulary in the options, and others had no science content in the stem but included science vocabulary in the options. In addition, 10 of the sentences with science content in the stem each had a single variant developed with no science content in the stem but with similar sentence structure and identical answer choices, as with the variants of the sports SNCPs.

Seven of the 20 science SNCPs showed C DIF against females, four showed B DIF, and nine showed A DIF. This result represents a higher proportion of science SNCPs showing C DIF against females than is usually observed in regular SAT pretesting, reflecting the researchers’ goal in this study to include a higher proportion than usual of items with *moderately specialized* science content and vocabulary. Table 7 presents seven of the science SNCPs, including some A DIF, B DIF, and C DIF items. Table 8 presents four additional science SNCPs along with the particular variant of each of the four items.

The first two items in Table 7 provide examples of SNCPs with substantial science content in the stems: “separating radium from radioactive residues,” “the therapeutic potential of radium,” “natural and synthetic examples,” and “type of catalytic molecule.” Despite such science content, however, females performed only 2% and 5% (respectively) less well than matched males. The vocabulary in the answer choices in these two items is not particularly scientific or challenging, and both of the items were relatively easy for the total population.

The third item in Table 7, about hydrogen gas, was very difficult for the total population, and it is classified as C DIF against females. But only 5% fewer females than matched males chose the correct answer (A), so this item provides another example of artifactual MH D-DIF. Of interest is that 7% fewer females than matched males chose distractor (E), whereas 7% more females than matched males chose distractor (C). The first terms in options (A), (C), and (E) all make sense when inserted in the sentence, so there appears to be differential knowledge by gender of the words “panacea,” “artifice,” and “stopgap,” although only “panacea” can be considered scientific vocabulary.

The final four items in Table 7 are one-blank questions that include scientific vocabulary in all of the answer choices. The first two of these (about political pundits and Coney Island) do not include science content in the stems, yet both show very significant C DIF against females: “seismic” was chosen by 16% fewer and “eclipsed” by 21% fewer females than matched males. For the total population, these items were both middle difficulty ( $\delta = 11$ ), but females clearly were not as familiar as males with the vocabulary in the options, or alternately, females responded differently from males to the science vocabulary in the options when confronted with it outside of a science context as established by the stem of the question. The final two items in Table 7 include scientific content in the stems as well as scientific vocabulary in the options, yet the performance of females relative to matched males is better than in the previous two items: 9% fewer females chose the key, “hover,” in the hummingbird item, and only 2% fewer females chose the key, “camouflage,” in the leafy sea dragon item. Historically, both “eclipsed” and “camouflage” have shown a pattern of C DIF against females in SAT-CR SNCPs; however, it appears that these words are more likely to yield C DIF against females when they are presented in the context of a nonscience item than when they are presented in a science context as established by the item stems. This observation suggests the importance of increased context in the item stem for decoding item keys, and by extension, for test fairness.

Table 8 presents four pairs of item variants. The first item’s stem, about “the pressure that ice exerts on Earth’s surface,” has been changed in its variant to a focus on “the popularity of jazz music over the past century.” In the next pair of variants, the topic has been changed from the movement of “leopard seals” to the knowledge of “several languages.” In each pair of item variants, similar to the sports variants discussed previously, the sentence structure is similar and the answer choices are identical. Only the content in the stem has shifted from science to nonscience.

In three of the four pairs of variants in Table 8, the performance of females relative to matched males is better in the nonscience variant. In the first pair, 12% fewer females chose “considerable” in the science version, and 9% fewer chose it in the nonscience version. In the second pair, 11% fewer females chose “agile” in the science version, and 8% fewer females chose it in the nonscience version. And in the final pair of variants in Table 8, 8% fewer females chose “feeble” in the science version, and 3% fewer females chose it in the nonscience version.

Only in the third pair of Table 8 item variants is this pattern broken: 11% fewer females than matched males chose the key in the science (“frogfish”) version, whereas 13% fewer females than matched males chose the key in the nonscience

**Table 7** Seven SAT-CR Sentence Completion Items With Science Content Administered in Study

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF †	Item text
F5-24	#4	8.4	0.57	-0.41	A		Believing that science should strive to — suffering, Marie Curie — a process for separating radium from radioactive residues so that the therapeutic potential of radium could be explored. (A) anticipate ... adopted (B) condemn ... requested (C) deplore ... detected (D) alleviate ... developed <sup>b</sup> (E) model ... exposed (OMITS)
F5-36	#1	9.6	0.67	-1.02	B	0 0 1 -2 0 0	While both natural and synthetic examples of the first type of catalytic molecule are —, there are few examples of the second type of catalytic molecule. (A) irrelevant (B) controversial (C) mysterious (D) intricate (E) abundant <sup>b</sup> (OMITS)
F5-18	#6	17.2	0.51	-1.70	C	1 1 1 2 -5 1	Hydrogen gas has been — as a clean and inexhaustible energy source, but skeptics warn that it is no —; alone, it will not permanently satisfy the world's energy needs. (A) touted ... panacea <sup>b</sup> (B) denounced ... luxuriance (C) acclaimed ... artifice (D) derided ... palliative (E) hailed ... stopgap (OMITS)
G8-10	#6	11.8	0.58	-2.03	C	2	Political pundits described the shift in voting patterns as nothing less than —, so strong and widespread was its impact. (A) automatic (B) periodic (C) seismic <sup>b</sup> (D) amorphous (E) elliptical (OMITS)

Table 7 Continued

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF †	Item text
G8-26	#3	11.3	0.62	-2.83	C		By the mid-1920s, the United States had more than 1,500 amusement parks; however, the stupendous Coney Island, dubbed "America's Playground," — all the others.
						-21	(A) eclipsed <sup>b</sup>
						11	(B) calibrated
						1	(C) systematized
						1	(D) reproduced
						3	(E) orbited
G8-28	#3	7.4	0.61	-3.24	C	4	(OMITS)
							The rapid beat and motion of its wings allow a hummingbird to — while it feeds, so that it appears suspended in midair.
						2	(A) revolve
						2	(B) alight
						4	(C) roam
						-9	(D) hover <sup>b</sup>
G8-06	#7	8.4	0.47	-0.36	A	1	(E) hibernate
						1	(OMITS)
							When floating in Australia's sea-grass beds, leafy sea dragons can successfully avoid predators thanks to their — of tendril-like fins and frilly appendages.
						0	(A) photosynthesis
						0	(B) environment
						1	(C) pigment
						1	(D) velocity
						-2	(E) camouflage <sup>b</sup>
						0	(OMITS)

*Note.* DIF = differential item functioning; MH D-DIF = Mantel-Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.



**Table 8** Four SAT-CR Sentence Completion Items With Science Content and Variants Administered in Study

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-26	1	10.9	0.73	-2.16	C		Because the pressure that ice exerts on Earth's surface is —, the rock beneath glaciers and ice sheets is pushed downward.
						3	(A) undetectable
						1	(B) debatable
						-12	(C) considerable <sup>b</sup>
						4	(D) imprecise
F5-24	1	10.5	0.62	-1.32	B	1	(E) variable
						3	(OMITS)
							Because the popularity of jazz music over the past century has been —, jazz's influence on other art forms is now apparent.
						2	(A) undetectable
						2	(B) debatable
F5-32	2	8.4	0.59	-3.72	C	-9	(C) considerable <sup>b</sup>
						2	(D) imprecise
						2	(E) variable
						2	(OMITS)
							Up to twelve feet long and weighing more than 1,000 pounds, the leopard seal is surprisingly —, moving quickly and gracefully.
F5-34	2	8.4	0.59	-2.05	C	3	(A) aggressive
						6	(B) leisurely
						-11	(C) agile <sup>b</sup>
						0	(D) noisy
						1	(E) bulky
						1	(OMITS)
							Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly — mind, perceptive and quick.
						6	(A) aggressive
						1	(B) leisurely
						-8	(C) agile <sup>b</sup>
						0	(D) noisy
						0	(E) bulky
						1	(OMITS)

Table 8 Continued

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
F5-38	4	9.8	0.32	-1.70	C		Frogfish have the ability to change color and blend into the background of their habitat; this — often makes locating them a — task. (A) concealment ... facile (B) masquerade ... paltry (C) camouflage ... formidable <sup>b</sup> (D) simulation ... fortuitous (E) tractability ... confounding (OMITS)
F5-36	4	12.9	0.36	-1.42	B		Mary Ann often hides her anger or distrust behind declarations of friendship; this — makes discerning what she really thinks a — task. (A) concealment ... facile (B) masquerade ... paltry (C) camouflage ... formidable <sup>b</sup> (D) simulation ... fortuitous (E) tractability ... confounding (OMITS)
F5-28	1	11.8	0.67	-1.10	B		Because surface tension exerts a relatively — force on large objects, objects that rely on surface tension for locomotion must be quite small. (A) feeble <sup>b</sup> (B) complex (C) timely (D) clear (E) constant (OMITS)
F5-30	1	12.4	0.53	-0.41	A		Because her predecessors exerted a relatively — influence on this artist, her paintings include few allusions to their works. (A) feeble <sup>b</sup> (B) complex (C) timely (D) clear (E) constant (OMITS)

Note. DIF = differential item functioning; MH D-DIF = Mantel-Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

**Table 9** SAT-CR Passages With Science Content Administered in Study

Passage	Word length	Number of items
Subatomic particles (nonspecialized)	106	4
Anti-matter (moderately specialized)	115	4
El Chicón (moderately specialized)	114	4
Cnidaria (more than moderately specialized) <sup>a</sup>	109	4
Heisenberg's uncertainty principle (more than moderately specialized) <sup>a</sup>	699	16

<sup>a</sup>Reproduced in Appendixes A and B. These are the two science reading passages for which some or all of the related items and their data are included in this report and in the cognitive interviews in 2011.

(“MaryAnn”) version. It is tempting to conclude that the word “camouflage” is, therefore, less familiar to females than to males, perhaps because of its use in military as well as scientific contexts. Yet recall that “camouflage” showed virtually no DIF against females as the key to the final item in Table 7. In the leafy sea dragon item, however, all the vocabulary can be construed as scientific, which perhaps helped this item become one of the few SNCs with “camouflage” as the key not to yield C DIF against females between the years 2005 and 2011, because all of the language in the leafy sea dragon item is clearly placed within a science context. By contrast, in the “MaryAnn” variant, the word “camouflage” is taken out of a science context and placed into a different framework, much like the items on political pundits and Coney Island in Table 7. Such a scenario seems to yield greater differences in performance than when all the vocabulary choices provided with the item are clearly scientific. Although it is true that similar items in this study (not printed in this report) with no science content in the stems but with scientific vocabulary such as “barometer” and “polarizing” in the keys showed little or no DIF against females, on the whole, it seems that placing scientific vocabulary in a nonscience context in SNCs can lead to increased differences in performance between males and females matched for ability, suggesting (again) the importance of increased context for test fairness as a whole.

### Science Content in SAT Critical Reading Passages

Five SAT reading passages with science content and 32 related passage-based items were administered as part of this study, as shown in Table 9.

Of the 32 passage-based items related to the five passages with science content, none show C DIF against females, even though they might have been anticipated to do so based on prior research on differences in performance between matched males and females on reading comprehension items associated with passages with science content (Doolittle & Welch, 1989; Lawrence & Curley, 1989; Lawrence, Curley, & McHale, 1988). In fact, only two of these 32 items show B DIF against females, and each of those show less than a 10% difference between matched groups of males and females. Because science content (unlike sports content) has long been specified for inclusion in the SAT-CR, the researchers sought out science passages for this study that were more specialized than routinely used in the SAT at the time the research was conducted. Yet even on items based on passages drawn from fields such as particle physics and quantum mechanics, and even on items dealing with rather dense and technical discussions of volcanic ash fallout and invertebrate species, matched males and females performed comparably. Table 10 includes all four of the items associated with the short passage on the phylum Cnidaria, and Table 11 includes seven of the 16 items associated with the long passage on Heisenberg's uncertainty principle.

As can be seen in Tables 10 and 11, some of the items ask about rhetorical features or the overall purpose of the passages; other items require inference, evaluation, analysis, or synthesis of the rather specialized ideas discussed in the passages. Some of the items are relatively easy for the total population ( $\delta = 6$  to 9), whereas others are of middle difficulty ( $\delta = 10$  to 12), and still others are quite challenging for the total group ( $\delta = 13$  to 14). But none of the items in Tables 10 and 11 shows more than a 6% difference between matched groups of males and females, and on several of the items, females perform 1%–3% better than matched males. The one item with B DIF in the direction of females uses the word “velocities,” which historically has shown a pattern of DIF against female test takers when used as a key in SNCs. Such data suggest that science reading passages that may be deemed *more than moderately specialized* do not result in items that show differential performance against females when matched for ability. It would therefore appear that the context provided in 100- to 800-word science passages puts matched male and female examinees on sufficiently comparable footing and eliminates the sort

Table 10 Critical Reading Items on Cnidaria Short Reading Passage

Test	Item number	Equated delta	r-biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
G8-08	11	10.01	0.42	-0.10	A	-1	The primary purpose of the passage is to
							(A) compare the behavior of jellyfish, sea anemones, and corals
							(B) question the adaptive features of jellyfish, sea anemones, and corals
							(C) examine the role of nematocysts in the phylum Cnidaria
							(D) highlight some important distinctions among members of the phylum Cnidaria
G8-08	12	6.2	0.60	0.58	A	0	(E) describe particular characteristics of members of the phylum Cnidaria <sup>b</sup>
							(OMITS)
							The reference to the letter "C" (line 1) is intended to assist with
							(A) classification
							(B) pronunciation <sup>b</sup>
G8-10	11	10.1	0.52	-0.02	A	-1	(C) location
							(D) memorization
							(E) generalization
							(OMITS)
							The simile in lines 4–6 ("Employed ... enemies") is used mainly to
G8-10	12	11.3	0.53	-0.42	A	0	(A) imply a biological connection between two dissimilar creatures
							(B) explain a complex behavior in more accessible terms <sup>b</sup>
							(C) establish a particular mood through imagery
							(D) symbolize cnidarians' place in the food chain
							(E) help readers visualize cnidarians at rest
G8-10	12	11.3	0.53	-0.42	A	1	(OMITS)
							In the passage, the "organization" (line 10) is described primarily to show how it
							(A) is used by cnidarians to control the flow of fluids
							(B) is consistent with cnidarians' methods of feeding and defense
							(C) reveals the need for cnidarians to have multiple layers of tissue
G8-10	12	11.3	0.53	-0.42	A	-3	(D) distinguishes cnidarians from most other invertebrates <sup>b</sup>
							(E) helps cnidarians navigate through their environment
							(OMITS)

Note. DIF = differential item functioning; MH D-DIF = Mantel-Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

**Table 11** Critical Reading Items on Heisenberg's Uncertainty Principle Long Reading Passage

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
G8-26	16	10.7	0.54	-0.58	A	-4	The primary purpose of the passage is to
						1	(A) explain a scientific concept <sup>b</sup>
						2	(B) challenge a scientific law
						0	(C) describe a scientific controversy
						0	(D) propose a scientific experiment
						1	(E) expose a scientific mistake (OMITS)
G8-26	17	12.4	0.47	0.25	A	-2	The first sentence (lines 1–3) implies that a large scale examination of the universe would decrease the appearance of
						0	(A) comprehensiveness
						2	(B) uniformity
						0	(C) frenzy <sup>b</sup>
						0	(D) opacity
						-1	(E) stability (OMITS)
G8-24	17	14.2	0.28	-0.03	A	2	The author would most likely characterize information resulting from the “observations” (line 7) as
						0	(A) accidentally accurate
						-2	(B) regrettably useless
						-1	(C) unnecessarily exhaustive
						0	(D) fancifully ephemeral
						2	(E) necessarily limited <sup>b</sup> (OMITS)
G8-24	18	9.3	0.52	-1.03	B	3	Lines 8–10 (“The high-frequency ... velocities”) indicate that one way in which photons can affect electrons is by
						-6	(A) increasing their mass
						2	(B) altering their speed <sup>b</sup>
						1	(C) reducing their light
						0	(D) obscuring their position
						0	(E) decreasing their numbers (OMITS)

Table 11 Continued

Test	Item number	Equated delta	<i>r</i> -biserial	MH D-DIF	ETS DIF category	STD-EISDIF <sup>a</sup>	Item text
G8-26	18	9.1	0.42	0.14	A	0	The quotation marks in line 10 indicate
							(A) reported speech
							(B) sarcastic disapproval
							(C) antiquated terminology
							(D) figurative usage <sup>b</sup>
							(E) ironic understatement
G8-24	19	13.6	0.47	0.33	A	0	(OMITS)
							The author would most likely agree that although “we clumsy observers”
							(line 21) can
							(A) theorize microscopic uncertainty, we are unable to prove it
							(B) negate microscopic uncertainty, we are unable to create it
							(C) predict microscopic uncertainty, we are unable to understand it
G8-24	20	9.5	0.55	0.26	A	2	(D) forestall microscopic uncertainty, we are not fundamentally its target
							(E) recognize microscopic uncertainty, we are not necessarily its cause <sup>b</sup>
							(OMITS)
							In line 32, “forces” most nearly means
							(A) opens
							(B) compels <sup>b</sup>
G8-24	20	9.5	0.55	0.26	A	1	(C) imposes
							(D) attacks
							(E) raises
							(OMITS)

Note. DIF = differential item functioning; MH D-DIF = Mantel – Haenszel estimate of DIF effect size in the delta difference metric; STD-EISDIF = standardized difference in expected item performance.

<sup>a</sup>Values do not always sum to zero because of rounding. <sup>b</sup>Key, or answer, to each item.

of large differences observed in the 15- to 30-word science SNCPs and in the nonscience SNCPs with scientific vocabulary in the options. It is also notable that in the science items, there was no significant difference in performance between the short and long reading passages, unlike in sports, where *more than moderately specialized* material presented in one of the 100-word passages did show differences in performance between males and females when matched for ability. As noted in the discussion of the items associated with the short sports passage, however, it is possible that other factors were involved, such as the use of numerals and decimal points in the description of batting averages.

That said, one characteristic of reading passages administered in standardized skills tests, such as the SAT-CR, is that IA and DIF data can be collected only on the passage-based items, not on the passages themselves. Occasionally, there are patterns in the item-level data that allow inferences to be made about student responses to a particular passage, but such inferences are usually tentative and difficult to confirm. For this reason and others, the researchers decided to add cognitive interviews (which combine think-aloud methodology with researcher-administered verbal probes) to supplement their analysis of IA and DIF data. Specifically, the researchers were interested in whether there might be affective sources of construct-irrelevant variance that might make such material inappropriate for skills testing, even though the quantitative data demonstrated relatively little difference in performance between matched groups of males and females on SAT-CR passage-based items.

## Phase 2: Qualitative Study, 2011–2012

In 2011, the study entered a new phase: the gathering of qualitative data through cognitive interviews in an effort to help explain and elucidate the quantitative data already collected through nonoperational variable sections of the SAT. The interviews were conducted to answer the following additional research questions that had arisen during an analysis of the quantitative data:

- What factors can help to provide explanations for the different levels of difficulty, discrimination, and DIF exhibited by different SAT-CR items using sports and science content?
- Do differences in students' interest and familiarity with the material have an effect on performance?

While the quantitative results of the 2009 and 2010 testing demonstrated whether differences exist between matched groups of males and females on particular items with sports or science content, such data do not explain why the differences exist nor which parts of the SNCPs or passage-based items might be responsible for the differences. The quantitative study was also limited in that it could not gather information about how students responded to the reading passages themselves, nor how those responses correlate (or do not correlate) with the DIF data collected on the accompanying items. For example, in the science items, a pattern had emerged with regard to the SNCPs versus the passage-based items, with more DIF against females observed in the SNCPs. This observation suggests that females perform better on science items when the science is placed within a larger context through a reading passage; however, there are typically no data regarding test takers' reactions to the science passages. It is possible that particular science passages may elicit confused or negative reactions from some students, even if no DIF is observed in the accompanying items. By conducting cognitive interviews using some of the study passages and items, the researchers hoped to refine the results from the quantitative study, thereby enhancing efforts to improve overall test fairness.

## Participants

The recruitment plan for this study was to obtain a sample of 60 students that reflected the population of SAT test takers in terms of gender and ethnicity. To be eligible, students must have already taken the SAT. Students who had just completed Grade 11 or 12 and who had recently taken the SAT in the central New Jersey area were invited by e-mail to participate in a 90-minute research study conducted at the ETS campus in Princeton. To participate, test takers were required to verify their SAT-CR scores as well as provide a list of courses and grades. If a transcript was not available, a list of classes with grades was acceptable. As compensation, test takers were paid an honorarium and reimbursed for their travel expenses to ETS's Rosedale campus.

Test takers were recruited from across the full range of SAT-CR scores, using score data provided by the CB. Initially, three ranges of scores were set: high score 800–600, middle score 590–450, and low score 440–200. However, after there was difficulty recruiting students in the lowest score band, the ranges were altered slightly. The final bands of SAT-CR



**Table 12** Student Participants by Score Band and Gender

SAT-CR score band	Males	Females
High score	11	9
Middle score	10	9
Low score	8	13

scores were as follows: high score 800–600, middle score 590–480, and low score 470–200. The expected 20 test takers were recruited and participated in each of these three score categories, though the gender breakdown by score category was not as closely balanced as would have been desirable. See Table 12 for the breakdown by SAT score and gender.

The final sample included the targeted 60 students, of which 29 were male and 31 were female. The racial breakdown of the sample included 52% White test takers ( $n = 31$ ), 13% Black test takers ( $n = 8$ ), 8% Asian test takers ( $n = 5$ ), 5% Hispanic test takers ( $n = 3$ ), and 22% other test takers ( $n = 13$ ). This diversity was intended to approximate the SAT testing population and was spread across the score bands. (The high percentage of “other” test takers is due primarily to those participants who chose not to indicate their ethnicity and/or whose self-identified ethnicity provided to ETS did not match the data provided by the CB.) It should be noted that this sample size, while robust for a cognitive interview study, is necessarily limited when compared to the sample sizes for the quantitative data, a factor that should be taken into account when evaluating conclusions based on this group. Another limitation was participant motivation: Unlike for the quantitative study, participants for the qualitative study knew their responses were for research purposes only and that they would not contribute to their scores on a test used to make high-stakes decisions.

### Materials: Test Forms

Two sample tests were used in this study, each of which contained 18 SAT-CR items. (The sample tests are reproduced in Appendixes A and B.) The majority of the items were drawn from the 146 reading items that were administered in Phase 1 of the study in 2009 and 2010. Each sample test form was structured to look like a typical SAT-CR section, beginning with eight SNCPs related to science, sports, or neutral (nonscience or nonsports) topics. The neutral items were the same on both forms. They were included to approximate the content balance in operational SAT-CR forms as well as to ensure that the study goals were not transparent to the participants. Following the eight SNCPs, each form had two short reading passages, one on either sports or science, and the other on a neutral topic that was the same on both forms. These short reading passages were accompanied by a total of three items (two on the study passage and one on the nonstudy passage). The last set of reading items was based on a long sports or science passage in a different content category from the short passage on the same form. Many of the items that were selected either had shown DIF in the direction of females during the quantitative phase of the study, or, if not, might have been expected to have done so because of the perceived specialization of topic and/or the specific language used to discuss that topic. The two tests were roughly comparable in terms of item difficulty—Test 1 had an average delta of 10.70 and an average  $r$ -biserial of .49, and Test 2 had an average delta of 10.47 and an average  $r$ -biserial of .47.

### Student Survey

Each interview concluded with a posttest survey that asked students in an open-ended fashion to describe their interest and involvement in both sports and science. The questions were meant to capture students' interests and involvement in activities both inside and outside of school.

### Interview Procedure and Protocol

A protocol was written for the interviewers to use in conducting each cognitive interview. It provided a step-by-step guide of the data collection procedure with instructions for the interviewer and the wording to be used when speaking to the participants. The protocol contained all of the questions to be asked of the participants related to each of the SAT-CR passages and items. See Appendix C for a complete sample protocol. The protocols for the two tests were identical, except for the test questions administered. Each session began with a demonstration of thinking aloud by the interviewer, followed by a sample item for participants to use to practice thinking aloud. Both of these sample items were taken from

the CB's SAT website and were included in the protocol. The session followed with the participants thinking aloud as they read and answered each of the questions and the three passages. The protocol included space for concurrent notes for the interviewers to record concerning item content and answering strategy, item correctness, and item interest, as well as retrospective questions to ask following each set of questions. The sessions concluded with the interviewers asking the posttest survey questions aloud and capturing the participants' responses. The posttest survey asked explicitly about interest in and engagement with sports and science.

## Interviewers

Five interviewers (two males, three females) completed a half-day training session to learn how to conduct the interviews and how to enter the data after each session was completed. Each test-taker interview was audio recorded and later transcribed. Interviewers also provided a brief summary of each interview, capturing the student's abilities and any relevant information related to interests in (or aversion to) sports and science.

## Data Analysis

The transcribed interview notes were formatted and entered into the qualitative data software package NVivo. Information that linked the gender of the participants was removed from the interview files so the research staff who coded the data did not know test-taker gender while coding the data. Four research assistants (two males, two females) coded the data. Each interview was double coded, and the interviews were balanced so that there was an even split of interviews coded by two males, two females, or one male, one female.

The coding process began with the transcripts being read by the interviewers. Information about item correctness, interest in or aversion to sports, and interest in or aversion to science was coded within each interview transcript. Information about SAT-CR scores, exposure to science, and classroom performance in science was coded from the SAT-CR score reports and student transcripts. Gender was added back to the data file once all interview transcripts were coded for sports and science information. There were weekly meetings during the coding process to ensure consistent coding. When there was a discrepancy (the same data coded differently by the two coders), one of the researchers made the final decision regarding the more appropriate code.

Item correctness (correct–incorrect) and gender (male–female) were coded in one step. Sports and science interest was coded using a multistep process described later. An interest in sports or science was identified first by examining the student think-aloud responses to the items, comments while reading and referring to the passages, or comments while responding to the retrospective questions. This led to a count of indications of interest in sports and interest in science for each participant. The same process was followed for dislikes: comments referencing an aversion to something specific about sports or science during each interview. Information about students' interest in sports and science was also collected more generally prior to the session (e.g., high school science classes and grades) and the end of each session (e.g., asking students about the specific sports and science-related activities they do inside and outside of school). An overall rating of sports interest and science interest was generated for each student by first examining the number of likes and dislikes (aversions) in the interviews across all participants to determine the range of interest across the sample. Then the number of likes and dislikes for each participant was examined, taking into account the number of sports- or science-related activities each student pursued independently. The interest rating used was on a 5-point scale ranging from 1 (*low interest*) to 5 (*very high interest*).

A second round of qualitative analysis to the transcripts was conducted while examining the results to categorize test-taking behaviors. This coding was done by looking at each transcript to classify the order participants used during the reading passages. The categories were as follows: read the passage first and took notes, read the passage first (no indication of notes), started reading the passage and skipped to the items, and skipped the passage initially and read the items first. This coding process led to a summarization of participant information by SAT-CR score band, gender, race, sports interest, sports performance, science interest, science performance, and reading passage test-taking behavior.

## Results and Discussion

For all reading items included in the cognitive interviews, vocabulary appeared to have the strongest effect on performance. Both male and female participants frequently noted that they were unfamiliar with the words used in the passages

or items and/or they “would have used different words” than the ones provided. One female participant in the middle score band said that she found the words “tricky.” Many male and female participants had difficulty with pronunciation. Perhaps for this reason, the variable that appeared to have the strongest correlation with performance on the sample tests used during the cognitive interviews was past performance on the SAT-CR rather than the participants’ interests in or familiarity with the topics that formed the content of the SNCPs, reading passages, or associated items.

### Sports Content in Sentence Completion Items

The qualitative data support the assumptions made from the quantitative data about the importance of students’ knowledge of vocabulary in answering SAT-CR items correctly, particularly for the SNCPs. Ranging in length from 15 to 30 words, SNCPs require test takers to rely heavily on their knowledge of the specific words used in the item. As such, the words used in the stem and options of an SNCP often determine item difficulty and, to some degree, differential performance among matched groups of males and females to the extent that there may be differential exposure to particular vocabulary. In the posttest survey, one male participant in the middle score band even noted that he found the SNCPs on sports to be “easier” because he “know[s] the topics well enough to know the sports vocabulary.”

In the hockey SNCP (Item 7 on Test 1), understanding the phrase “body check” in the stem of the item seemed critical to whether a participant was able to answer the item correctly. Either the participant needed to know what the term meant or how to deduce the meaning from the context of the item. One female participant said during her cognitive interview, “Body checking made it [the item] easy to answer,” whereas another female participant commented, “I’m not an ice hockey fan, so had to think what ‘body check’ meant.” One male participant explained that knowing what “body check” meant helped him to answer the item: “As a hockey player, I completely understand he was being physical, not cunning or inconsistent.” Another male said, “Without that background knowledge, the question would have been more difficult.” At the same time, however, several male participants commented that they were unfamiliar with the phrase “body check,” but knew that they could, in the words of one male participant, assume “body checking means something physical” because of the inclusion of the word “body” in the term.

Similarly, for the golf SNCP (Item 2 on Test 1), the golf context made the item appear more difficult to participants who lacked familiarity with the sport. However, whether participants were interested in or familiar with golf and its terminology seemed to have little effect on their actual performance. For this SNCP, most participants seemed able to use logic to select the correct option, regardless of their familiarity with terms such as “par.” As one female participant in the high score band explained, “The context ... suggests this item is about inconsistency.” Another female participant said that she “found some terms confusing” because she does not play golf; nevertheless, she answered the item correctly based on logic.

Logic was less helpful in decoding the short-track skating item (Item 7 on Test 2). In this case, the challenging vocabulary had less to do with sports and was instead, in the words of one female participant, “just ridiculously hard.” The item ( $\delta = 16$ ) exhibited C DIF against females in the quantitative phase of the study, but the percentage difference between matched groups of males and females was only 6%, which is a difference that can be deemed artifactual. In the qualitative phase of the study, both males and females seemed to struggle equally with the challenging vocabulary in the options: “mercurial,” “tractable,” “puerile,” and “rife.” One female participant in the middle score band noted that “the part saying ‘roller derby on ice’ made it [the item] harder because I don’t know what a roller derby is.” As an interviewer noted about a different female participant in the middle score band, “It was the vocabulary that impeded her answering correctly. She ... didn’t know what to fill the blank with because the vocabulary was too high level.”

### Sports Content in Critical Reading Passages

With regard to the reading passages on the sample tests, more participants, both males and females, complained about the fairy tales passage—taken from a published form of the *PSAT/NMSQT*® test to balance test content—than about either the sports or the science passages: “I really hate fairy tales” versus “I am a sports fanatic” and “I’d watch almost every sport you can find.” Many male and female participants were visibly excited to encounter sports, even when they did not seem to understand the short baseball passage or answer the associated items correctly. One of the low-scoring male participants rhapsodized about the short baseball passage, which appeared on Test 1, “I love baseball,” even though he did not answer any of the associated items correctly.

Of the female participants, most reported liking the topic of the short baseball passage because it was “more interesting” than typical SAT-CR passages, although several female participants specifically complained about the parenthetical discussion of batting averages. One female participant in the low score band stated that “boys would understand this more than girls,” whereas another in the middle score band noted that the statistical information in parentheses was “confusing and unnecessary.” One specifically noted that she was not interested in statistics; yet another in the high score band said, “I don’t know what a .432 batting average means.” From the interview transcripts, it would appear that several female participants across all three score bands were puzzled by this part of the passage, either stopping or stumbling over it. In trying to decipher the parenthetical description of batting averages, some female participants appeared to lose sight of the meaning of the passage as a whole, suggesting that there was an additional source of variance—namely, math—an examination of which was outside the scope of this particular study.

These particular lines of the Wee Willie Keeler passage did not pose problems only for female participants, however. The word “accrue” caused both male and female participants to falter, as there was little familiarity with the meaning of the word among all participants, regardless of gender or ability level. And at least one male participant in the high score band similarly questioned the meaning of the short passage’s parenthetical discussion of Wee Willie Keeler’s batting average: “Is a batting average of .432 good? Is it bad?”

Yet the placement of the sports content in the context of a short reading passage, as opposed to appearing in an SNCP, did seem, in general terms, to help both male and female participants to answer the items correctly. The greater context provided by the short reading passage helped to explain any particular sentence—or word, such as “accrue”—that might have been confusing on its own. The interaction of the items with the passage also helped. Most participants noted that answering the items helped them to understand the short baseball passage as a whole. One female participant in the middle score band explained that this was “because it made me look back at the passage to see what I didn’t see before and read between the lines.” Another female participant in the middle score band noted, “You have to look through each clue,” thereby indicating that the options in the items gave her a clearer sense of the passage’s meaning. Even for the description of batting averages, context helped. One female participant in the high score band noted, “The word ‘astonishing’ helped me figure out it [the batting average] must be good.” Indeed, one female participant in the low score band went so far as to say, “Maybe if it [the passage] were longer, it would have helped me,” suggesting the significance of increased context for understanding challenging material.

With regard to the long passage on women’s boxing, which appeared at the end of Test 2, most participants across all three score bands seemed to find the passage engaging and accessible. The affective feedback on the passage was, on the whole, extremely positive. Several female participants noted that the “personal story” aspect of the passage made it appealing; both male and female participants compared it to a relatively recent film about boxing, *Million Dollar Baby*. The narrative style of the passage, focused on the emotions of a central female character, seemed appealing. As one female participant noted, “I like this passage better than the other two, but again, that’s because I like the sport of boxing. I’m sure if someone liked jellyfish or ... fairy tales, they’d like the other two better than this one.” Said another, “They should make more passages like this. It keeps you focused.”

Enjoying the passage, however, seemed to have little effect on participants’ abilities to answer the associated items correctly. Participants expressed pleasure in encountering the subject of boxing on a standardized test, but many still stumbled over the specific vocabulary they encountered, such as “nostalgia,” “tenacious,” “capricious,” “petulant,” and “heckling.” Some phrases also appear to have been *specialized* for those participants less familiar with boxing. As one interviewer noted of a female participant in the middle score band, “The idea of a ‘physical reminder’ [line 4] was a confusing idea for her to grasp, so she guessed”—incorrectly—when answering the item related to the phrase.

As with the short baseball passage on Wee Willie Keeler, however, the inclusion of additional context that is inherently incorporated within a longer passage seemed to help participants to key the items correctly even when they experienced difficulty with vocabulary or concepts raised in the passage as a whole. As one male participant in the high score band noted, “The questions helped my understanding because you have to read the main parts over. Answering the questions helped me focus.” Similarly, a different male participant in the high score band said that even when he was confused by boxing terminology, such as “double hook” (Line 51), “work her corner” (Line 18), and “stay in the pocket and fight” (Line 23), answering the questions helped him “because I had to go back over the different parts of the story,” thereby increasing the understanding of what he had read. In this way, the increased context provided by the passage for each item appeared

**Table 13** Performance on Sports Items by Level of Sports Interest

Interest in sports	Score band	Range of sports (% incorrect)	Mean sports (% incorrect)	Total students	Males	Females
5	High 600–800	25.00	25.00	1	1	0
	Middle 590–480	0–33.33	11.57	6	5	1
	Low 470–200	0–50.00	31.35	7	5	2
4	High 600–800	0–33.33	11.57	6	4	2
	Middle 590–480	0–75.00	33.68	8	3	5
	Low 470–200	22.22–44.44	28.70	9	2	7
3	High 600–800	0–11.11	5.56	2	0	2
	Middle 590–480	N/A	N/A	0	0	0
	Low 470–200	25.00–55.56	35.19	3	0	3
2	High 600–800	11.11–33.33	22.22	3	0	3
	Middle 590–480	0–55.56	29.45	5	2	3
	Low 470–200	75.00	75.00	1	0	1
1	High 600–800	0–33.33	9.72	8	6	2
	Middle 590–480	N/A	N/A	0	0	0
	Low 470–200	N/A	N/A	0	0	0

to help participants decode words or phrases that they might otherwise have found confusing, perhaps contributing to the low levels of DIF observed in the items associated with this passage during the quantitative phase of the study.

### ***Performance on Sports Items by Level of Sports Interest***

In addition to the qualitative feedback collected during the think-aloud portions of the interview, each participant was asked a number of questions during the posttest survey about interest and involvement in sports. The research team reviewed all participant responses and made categories of sports interest according to the number of activities mentioned and the enthusiasm expressed toward sport-related topics in the poststudy survey. Interest level was categorized on a 5-point scale ranging from 5 (*a high interest in sports*) to 1 (*a low interest in sports*). The sports interest level of the participants in this study covered the full range of the scale with participants in each category. Interest in sports was then examined by each participant's score band, gender, and the percentage of sports items that were answered incorrectly to seek patterns in performance. Because so few sports items appeared on each 18-item test form—four items on Test 1 and nine items on Test 2—it was determined that answering an item incorrectly provided more useful information about whether interest in sports factored into performance, as opposed to skipping the items or answering them correctly.

Of the 13 participants who indicated the highest interest in sports (interest level = 5), there were significantly more males than females, as seen on the left side of Table 13, with 10 males and three females represented. Among the participants who indicated the highest interest in sports, there was a range of ability levels, with some participants in each score band reporting the highest interest, although there was only one such participant in the high score band. Notably, there was no discernible pattern in performance according to score band or gender in this group of students. Interest in or aversion to the sports material appeared to have no discernible effect on performance in answering these sports questions.

More females than males reported a moderately high level of interest in sports (interest level = 4)—14 females, nine males—as seen on the right side of Table 13. There was a fairly even division of participants by score band in the moderate interest in sports, with six participants in the high score band, eight in the middle score band, and nine in the low score band. Again, there appears to be no distinguishable pattern with regard to gender or performance. Instead, the participants in the high score band performed better than the students in the low score band, with participants in the high score band answering fewer sports items incorrectly than students in the low score band. From these data, it seems fair to say that participants who indicated a high interest in sports did not appear to have an advantage in answering the items with sports content. Nor were participants with a low interest in sports at a disadvantage.

Five students (all female) indicated a medium interest in sports (interest level = 3). This group displayed a range of performance on sports items, with those in the high score band performing better than those in the low score band, as seen in Table 13. The performance on sports items in this small group of participants does not indicate any unexpected patterns of performance. The participants in the high score band performed better than participants in the low score band, answering fewer sports items incorrectly.



More females — seven females, two males — indicated a moderately low interest in sports (interest level = 2), as seen on the left side of Table 13. Among this group, there was no apparent pattern on the performance of sports items according to either score band or gender, although the student (a female) who answered the most sports items incorrectly was in the lowest score band.

When examining participants who ranked in the lower levels of interest in sports (interest level = 2 or 1), the gender split is more balanced: nine females, eight males. Participants with a lower level of sports interest were more likely to be in the high score band, yet this low interest appears to have had no effect on their performance. The participants in the high score band answered few or no items incorrectly, as opposed to participants in the middle or low score band, who answered more sports items incorrectly.

An interesting pattern did emerge among the participants who indicated the lowest interest in sports (interest level = 1) — six males, two females — as seen on the right side of Table 13. All eight participants who reported the lowest interest in sports were in the high score band, and they performed well on the items with sports content, answering few or no items incorrectly.

Participants may have loved sports, but they did not appear to perform better on items with sports content as a result. Also, there was no observable difference by gender when matched for ability.

### ***Science Content in Sentence Completion Items***

For the science SNCPs, comprehension of vocabulary remained critical to performance. Even among high-ability students, the more challenging vocabulary caused several students to pause. For example, on both the camouflage item and its variant (Item 6 on both sample tests), the vocabulary in the options appeared to pose the greatest challenge to keying the item correctly, with some male and female participants noting the following specific words as causing difficulty: “facile,” “paltry,” “camouflage,” “fortuitous,” and “tractability.”

For the leopard seal item, which had shown C DIF against females in the quantitative phase of the study, it appeared that much of the issue rested with vocabulary. The very concept of a leopard seal — bringing together the opposite concepts of “leopard” and “seal,” predator and prey, within the name of one animal — was pointed out as “extremely confusing” by some female participants. By contrast, males seemed to have an easier time skimming past the name of the animal to answer the question by eliminating all options except “agile,” thereby reaching the key more easily. As one female interviewer said of a male participant in the middle score band, “At first, he could not read ‘agile’ correctly, then after repeating it over, said, ‘Oh, *AGILE!* Duh!’ and correctly chose it as the answer. He said this question was medium difficulty because ‘agile’ messed him up at first, then it hit him what it was.” The same process was observed with a male participant in the high score band in answering Item 6 on Test 2: “He had not learned about frogfish but had learned about camouflage with chameleons . . . . What made the question easy was ‘camouflage.’ What made the question harder was not knowing the definitions of ‘paltry’ and ‘tractability.’”

Vocabulary was particularly critical to performance in the case of science vocabulary placed into a nonscience context. For example, in the item on Coney Island (Item 8 on Test 1), one male participant in the low score band observed, “What made the question harder was that I felt ‘systematized’ and ‘eclipsed’ were hard to use in that situation.” Several males in all three score bands noted that it was hard to select the correct answer from the options because the words were not what they had expected. One female participant in the middle score band in the study refused outright to answer the item, on the basis of its vocabulary, although it should be noted that she had experienced increasing difficulty with the vocabulary in all the SNCPs up to this point.

With regard to political pundits (Item 8 on Test 2), many male and female participants seemed to expect a key that would mirror the wording in the stem, “strong and widespread,” yet were slow to select “seismic” as the correct answer because of a lack of familiarity with the word. Several male and female participants noted that the vocabulary in the options of the political pundits item was “difficult” and “different from what I’d expect.” Said one male participant in the high score band, “After reading it [the question], I had an idea that it had to be something enormous or grand and I thought ‘seismic’ but I have never seen it used in that way.” Some participants skipped the question completely because they did not know the words — not only “seismic” but also “amorphous” and “elliptical” — and “didn’t want to risk getting the answer wrong” (to borrow the words of a female participant in the middle score band). Most of all, though, female participants seemed to have somewhat lesser familiarity than their male peers with the term “pundit,” which appeared in the item’s stem, reflecting the item’s performance in the quantitative phase of the study, where it had shown C DIF against

females. As one female participant in the middle score band noted, “I didn’t know what a pundit was, so I was forced to guess.”

Along similar lines, several female participants reported difficulty with the glaciers SNCP (Item 5 on Test 1), which had shown C DIF against females in the quantitative phase of the study, because the key “considerable” was, again, not what they had expected in this context. In this case, female participants found the key “considerable” to be insufficiently scientific. Several female participants noted that the nonscientific nature of the key “threw [them] off.” Said one female participant in the middle score band, the “wording of the question was confusing” largely because of the use of nonscientific vocabulary in the options. This perceived mismatch between the expected key and the actual key was not exclusive to female participants. Some male participants also said that they found the lack of scientific vocabulary to be puzzling, with one male noting that he could understand scientific material better “when scientific terms were used.” One male in the high score band who had studied geology said that the answer to the glaciers item “wasn’t immediately clear” because none of the options seemed “right” for the context. Such reactions were in contrast to responses for the jazz variant of the item (Item 5 on Test 2), for which male and female participants reported difficulty — “some words mean similar things like undetectable/debatable so it’s trying to trick me” — but these challenges were more easily resolved and the item more easily keyed by both male and female participants across all three score bands.

From this evidence, it appears that both male and female participants were trying to key the SNCPs based on the words that they felt should appear in the blanks. When the type of vocabulary used in the stems and the options did not match expectations — for example, when science vocabulary appeared in a nonscience context or nonscience vocabulary appeared in a science context — this shift affected the keyability of the item, regardless of the participant’s gender.

### ***Science Content in SAT Critical Reading Passages***

While lack of familiarity with vocabulary had a more significant effect on participants’ abilities to select the correct keys for the SNCPs, it also played an important role in participants’ performances on questions associated with short reading passages. For example, female participants in the high score band noted of Cnidaria, which appeared on Test 2, “The big science words confused me,” and “What made the passage confusing was that a lot of the words are hard to say.” One female participant in the high score band singled out “mesoglea” and “radial symmetry” in particular as needing further explanation. Other male and female participants across all three score bands stumbled over “cnidarians,” “anemones,” “nematocysts,” and “coelom.” Interviewers noted of several participants that they “had a lot of trouble sounding out the difficult biology vocabulary,” which most had not encountered before. Said one female participant in the middle score band, “It was too technical and the words were too hard to pronounce.” One interviewer wrote of a female participant in the middle score band, “She worked hard at pronouncing the difficult science jargon, but I could tell that she was frustrated,” and she answered both questions incorrectly. Many participants sighed, fidgeted, fumbled, complained — even those who indicated scientific interests deep enough to be contemplating future careers in science and medicine.

At the same time, participants were clear that answering the items helped their understanding of the Cnidaria passage as a whole, often making the difference between guessing and engaging in an educated selection among the different options. Noted one male participant in the high score band, “Answering the questions made it easier to understand the passage because finding the main point means you can use that to understand facts you don’t know.” Along the same lines, a female in the middle score band noted that answering the questions “made me look at the purpose and not get confused about the words I don’t know.” This same female participant had particular trouble with the word “dibloblastic,” but she was able to interpret the word’s meaning by answering Item 10 about the creature’s organization and, in so doing, key the item correctly.

That increased context helped both male and female participants work through challenging material and find the correct keys for the items was particularly evident in the long passage about Heisenberg’s uncertainty principle, which appeared on Test 1. Of all the study passages included in the cognitive interviews, Heisenberg’s uncertainty principle was undoubtedly the most challenging, with strong negative affective responses to the passage reported by most participants. Many commented in the posttest survey that they “hated reading about it,” with a male in the low score band noting that he found the passage “boring and pointless.” To quote a sampling of students, “I don’t think anyone could understand it” and “I find it really boring. There’s only a couple things I like about science and definitely not SAT passages about science. Those kill me.”



Perhaps because both male and female participants struggled with the material, they often answered the associated items incorrectly. Many participants did not even try to deduce the correct answer but simply gave up and guessed. Even participants whose official SAT-CR score band was high often answered the items associated with the Heisenberg's uncertainty principle passage incorrectly. Significantly, males appeared to struggle just as much as females did, and, when matched for ability, performance was comparable on these items. Interviewers reported males "taxed, yawning, squirming, mumbling" just as much as females. As one male participant in the low score band noted, "Most of that stuff was confusing. I am just not into science."

Again, much of the difficulty appears to have derived from the specific vocabulary in the passage. Participants' lack of familiarity with the vocabulary was evident in the interview transcripts. They stumbled when reading aloud not just the scientific terms but also words such as "frenetic," "intrinsically," "quiescent," and "relinquishing." Several male and female participants across all three score bands noted that they did not understand many of the words in either the passage or the items. One female interviewer noted of a male participant in the low score band, "This interview was fascinating in that the student appeared to misunderstand much of the material based on the amount of misreading of words, word substitution, and the frequent mispronouncing of other words." One female participant in the middle score band stated that she found the passage to be "extremely difficult as well as uninteresting and confusing" because of the "scientific jargon." Another female participant in the middle score band complained that there were "too many scientific words, making it difficult to understand"; the same female participant said that answering the items made her more confused about the meaning of the passage, not less. Yet another female participant in the low score band said that answering the questions did not help her to understand the passage because she "had to look at the passage in depth more than necessary, and that was confusing." Such statements were contrary to participants' responses regarding the items for other reading passages, in which the majority of both male and female participants commented that answering the items had assisted in their understanding of the passage as a whole. Yet for Heisenberg's uncertainty principle, participants complained that they just "couldn't relate to it." As one female participant in the high score band commented, "This is a reading test but it felt more like trying to understand science. It felt like a science lesson. If I knew it [the content], it would've been easier."

Knowing the content, however, did not necessarily help, and participants were not good predictors of their own performance. Some participants who stated that they believed prior knowledge of the topic helped them to answer the items correctly actually performed quite poorly; such participants also tended to misidentify the subject of the passage as being chemistry instead of physics. One female participant in the low score band who stated that she "enjoys science" and the passage "didn't bother me" answered only three of the eight associated items correctly. This same participant was described by the interviewer as fidgeting and sighing loudly while reading the passage, with body language that belied her assertions that she felt confident answering the items. By contrast, a female participant in the high score band who expressed much intimidation and anxiety about science described this particular passage as "a punchy, entertaining read" and answered most items correctly. This participant stated that the author's use of similes and metaphors helped her understanding of the challenging material. She, like several other high-scoring female participants in the study, easily answered the item referencing "velocities" because of her knowledge that "speed and velocity are synonyms." In the quantitative phase of the study, this particular item had exhibited B DIF against females—making it the item with the most significant DIF against females within the context of this particular passage. In the cognitive interviews, however, the best predictor of performance on this item, as on all the items associated with the reading passages, was neither gender, interest in, nor aversion to the material but rather the participants' actual SAT-CR performance as documented by their official SAT-CR scores.

### ***Performance on Science Items by Level of Science Interest***

As with sports, each participant was asked a number of questions during the posttest survey about interest and involvement in science. The research team reviewed all participant responses and made categories of science interest according to the number of science activities mentioned and the enthusiasm expressed toward science-related topics in the poststudy survey. As with sports, interest level was categorized on a 5-point scale ranging from 5 (*a high interest in science*) to 1 (*a low interest in science*). No participants in the sample expressed an interest at the highest level (5), but there was a range of interest reported in the other four categories.

In addition, high school transcripts and/or report cards were requested for each participant. Nearly all participants provided a transcript, a report card, and/or a list of courses with grades. Information on the number and type of science

**Table 14** Performance on Science Items by Level of Science Interest

Interest in science	Score band	Range of science (% incorrect)	Mean science (% incorrect)	Total students	Male	Female	Mean science exposure	Mean classroom science performance
4	High 600–800	0–50.00	14.44	5	3	2	High	3.75
	Middle 590–460	25.00–50.00	35.42	4	3	1	High	3.11
	Low 470–200	44.44	44.44	1	1	0	High	3.33
3	High 600–800	25–33.33	29.17	2	2	0	High	3.34
	Middle 590–480	55.56	55.56	1	1	0	High	3.00
	Low 470–200	0–55.56	33.33	3	2	1	Medium	2.67
2	High 600–800	22.22–50.00	30.56	4	2	2	High	2.71
	Middle 590–480	25.00–66.67	48.33	5	5	0	High	2.58
	Low 470–200	0–77.78	46.97	11	4	7	Medium	2.91
1	High 600–800	0–50.00	9.03	8	4	4	High	3.31
	Middle 590–480	0–50.00	31.35	7	1	6	High	3.50
	Low 470–200	0–33.33	13.89	4	1	3	Medium	3.04

classes taken, as well as performance in each class, was standardized by the research team because some schools provided letter grades and some provided number grades. The small sample size also meant that information about science classes was collapsed into a category of science exposure. (There were not enough data to make more specific categories about science class information.) There were three levels of science exposure—high, medium, and low—according to the number of science classes participants had taken. Performance in each science class was estimated to fit a grade point average, with 4 as the highest performing and 0 as failing.

Student interest in science was then examined by each student's score band, gender, the number of science items that participants answered incorrectly, and the range of science classes that each student had taken, along with their average performances in those classes, to seek patterns in participants' performances on the science items in the study. Because so few science items appeared on each test form—nine items on Test 1 and four items on Test 2—it was determined that answering an item incorrectly provided more useful information about whether interest in science factored into performance on the science items, as opposed to skipping the items or answering them correctly.

As shown in Table 14, one half of the participants who reported a high level of interest in science (interest level = 4) were in the high score band, with five participants in the high score band, four in the middle score band, and one in the low score band. More males than females reported a high interest in science: seven males, three females. Science performance on the science items seems to correlate with the score band for the males, with the participants who were in the higher score band answering few or no science items incorrectly and participants in the middle and lower score band answering more science items incorrectly. Not many females indicated a high interest in science, and there was no discernible pattern among the female participants with regard to performance on the science items.

More males than females indicated that they had a moderate level of interest in science (interest level = 3), with five males and one female. There is no striking pattern in this specific range of science interest on science performance, suggesting that there is no correlation in this score band between interest and performance. The same can be said for participants in the low score band, though it should be noted that the performance on science items for participants in the low score band is the most difficult to interpret.

That said, when reviewing participants who indicated a moderate to high level of science interest all together (interest level = 4 or 3), there is a general pattern that participants who were in the high score band tended to answer fewer science items incorrectly. Participants in the high and middle score bands performed well on the science items, with more participants in the high score band answering few to no items incorrectly. The majority of participants in the groups with moderate to high interest in science who provided science class and performance data also had a high exposure to science via science classes. In this study, more male participants indicated a moderate or high interest in science than female participants (12 males, 4 females). The data also show that student interest in science and science exposure are correlated, though it should be noted that scientific exposure is not the same as science ability.

Participants who reported moderately low interest in science (interest level = 2) were spread across each score band, but there were more participants in the low score band, as seen in Table 14. This group of participants had an equal number

of males and females—nine males, nine females—but there were more females in the low score band (seven of the nine females). Participants in the middle and low score bands displayed more range, answering more science items incorrectly. Participants in the low score band answered the most science items incorrectly, with five of the six participants being female.

The participants who reported the lowest level of science interest (interest level = 1) were fairly spread out in terms of the score bands. As shown in Table 14, there were more females in this group: 15 females and six males. The females in this group were spread among the score bands with the most being in the middle score band, whereas the males were more likely to be in the high score band, with four of the six males in the high score band. There was a general correlation between score band and the percentage of science items answered incorrectly. Participants in the high score band answered few to no science items incorrectly, with the exception of one female who answered half of the science items incorrectly. Participants in the middle and low score bands had more of a range of answering science items incorrectly. There were two participants in the low score band answered all science items correctly. They both had been exposed to a high number of science classes, but they did not have very strong grades in those classes. Participants in the low score band answered the most science items incorrectly, with five of the six participants being female.

The data show no real correlation between interest in either science or sports and performance on items with such content. Instead, it would appear that participants' interest in a particular topic (or lack thereof) did not have a measurable effect on performance. This finding is contrary to earlier studies that found that test takers' interests correlated with performance (Buck, Kostin, & Morgan, 2002; Carlton & Harris, 1992; Willingham & Cole, 1997). In this study, participants may have stated that they loved sports or hated science, and they often stated that they believed this engagement made a particular item easier or harder, but their actual performances seemed to be neither helped nor hindered by their affective responses. Participants did not appear to perform better on items associated with passages in which they expressed interest. Nor were reading comprehension items related to science passages differentially more difficult for females than for males when matched for ability, as had been shown to be the case in past studies (Doolittle & Welch, 1989; Lawrence & Curley, 1989; Lawrence *et al.*, 1988). Instead, participants seemed to perform well or poorly based on overall ability, as measured by their official SAT-CR scores, rather than because they had any interest in or aversion to the specific topic of the material on which they were being tested.

### ***Possible Effect of Test-Taking Strategies***

What did seem to have some effect on performances were the participants' test-taking strategies, at least with regard to items associated with the long passages for both sports and science. Particularly for the long passage on women's boxing, 35% of the males skipped the passage completely in favor of immediately answering the items, compared with 6% of the females. Eight-six percent of the females read the passage in full before answering the items (see Table 15). From participants' think-aloud remarks during the cognitive interviews, it would appear that some males were inspired with false confidence by the sports content into thinking that reading the actual passage was unnecessary to their understanding of the items. As one male participant who did not read the women's boxing passage noted, "Watching boxing on TV made the questions easy"—yet he answered most of the questions incorrectly. Similarly, a male participant in the low score band who expressed a strong interest in boxing noted, "If I read the entire passage, I am liable to get bored and not focus." Another male participant in the middle score band explained that he could use his own background as an athlete to answer—yet he too answered almost every item incorrectly. In such situations, participants' familiarity with and interest in the topic of the passage not only did not help them to answer the items, it might have led them to use a test-taking strategy that was far from optimal in answering reading comprehension items, namely, not reading the passage on which the items were based.

While similar test-taking strategies were observed with regard to the long science passage, fewer males were likely to skip reading the passage on Heisenberg's uncertainty principle, likely because of the perceived greater complexity of the material. The test-taking strategies used by both males and females to approach the long science passage were more comparable than for the long sports passage, as demonstrated by Table 16. That said, more females than males read and took notes on the long science passage before tackling the items, with one female participant in the middle score band summarizing the main point of each paragraph and carefully noting it down on her test book. This suggests that females were still more deeply engaged with the passage than their male peers before trying to answer the associated items. Although the participants who skipped reading the passage tended to be in the low score band, and the participants who took notes

**Table 15** Test-Taking Strategies for Long Sports Passage by Gender

Strategy	Males	Females
Read passage first and took notes	0	0
Read passage first	8	13
Started passage, stopped, and skipped to items	1	1
Skipped passage completely	5	1

**Table 16** Test-Taking Strategies for Long Science Passage by Gender

Strategy	Males	Females
Read passage first and took notes	1	3
Read passage first	6	8
Started passage, stopped, and skipped to items	4	4
Skipped passage completely	3	2 <sup>a</sup>

<sup>a</sup>One of these two female participants wore a hearing aid and had difficulty speaking and so declined to read the long passage aloud.

tended to be in the high score band, that was not universally true: Participants in the middle score band in particular used a variety of strategies that spanned the full range of test-taking behaviors.

It is possible that these differences in test-taking strategies between males and females had a greater effect on their performances than any other factor, except for overall ability. Some males were even aware of the disadvantage that choosing not to read the passage had created, with one male participant in the middle score band responding that a particular passage-based item was “hard” because “I didn’t read the whole passage so trying to find the main idea of the passage as a whole was difficult.” The small sample size makes it difficult to generalize, much less make strong causal statements. Moreover, as previously noted, one of the limitations of the qualitative phase of the study was likely a certain lack of motivation, as participants were not taking a test that would be used in making high-stakes decisions. Nonetheless, the results raise questions about whether gender differences in test-taking strategies may be correlated with the smaller levels of DIF observed in reading comprehension items for long passages than in short passages or SNCPs. There was less pattern of difference in how participants approached the short passages or the SNCPs; the different strategies seemed to apply to the long reading passages only. It is possible that males “level the playing field,” so to speak, by approaching the SAT-CR passages in a different, and ultimately less methodical, fashion than their female peers do. This could be an interesting avenue for further study.

## Conclusions and Recommendations

To revisit the research questions that inspired this study, we sought to examine the following:

- If sports and science are to be included in the SAT-CR, what kinds of material are appropriate on which to base fair and valid score-based inferences?
- What constitutes *specialized* material for these content areas?
- Does item type matter with regard to DIF? Namely, are there observable differences in performance on SNCP versus reading-based items as well as on reading-based items for short passages versus long passages?

During the qualitative phase, we added additional research questions that had arisen during an analysis of the quantitative results:

- What factors can help to provide explanations for the different levels of difficulty, discrimination, and DIF exhibited by different SAT-CR items using sports and science content?
- Do differences in students’ interest and familiarity with the material have an effect on performance?

From the data from both phases of the study, it would appear that most sports and science material of the sort used in this study should be permissible in a skills test. Particularly in the context of a long passage, in which most information that students need to answer an item is provided, very little of our material appears to be too *specialized*. From participants’ statements during the cognitive interviews, it seems clear, too, that interest in and/or familiarity with a subject has very

little effect on performance on items associated with that subject. Instead, the real correlation is with overall ability. As such, SAT-CR items with sports and science content are unlikely to yield appreciable differences in performance between male and female test takers on their own; other factors have to be at work.

Indeed, very few of the research items, including those classified as C DIF, show significant differences between males and females when matched for ability. Instead, much of the observed C DIF is artifactual and, therefore, would ideally be ignorable. Many easy or hard items in this study are classified as C DIF using the MH statistic, suggesting that the items discriminate against females even though the percentage difference between matched groups of males and females (often as low as 5–6%) is smaller than the percentage difference observed in several middle delta items classified as B DIF. This suggests that, for some items, the MH statistic is misleading, indicating greater differences in performance between males and females than actually exist. This observation in some ways trumps the research questions with which we began the study, because to answer those questions about gender differences, there first needs to be confidence that the system used to identify C DIF items is providing actionable data. In this study, it was the use of STD-EISDIF in addition to MH statistics that gave us the information needed to evaluate the performance of males and females when matched for ability.

On the basis of these observations, we make the following recommendations.

Recommendation 1: Consider using more than one type of statistic to evaluate differential item functioning.

In the quantitative phase of the study, a pattern of 5–6% difference between matched males and females was observed for many items that were rejected from further use in testing on the basis of the MH statistic. This happens particularly for easy and hard items, whereas many middle difficulty items that are acceptable for further use based on the MH statistic show differences between matched groups of 10% or more when the STD-EISDIF statistic is used. Although it may be preferable to throw out too many rather than too few items based on their pretest data, it would be better still to use statistics that relate directly and accurately to the expected impact on a group's performance. Perhaps the time has come to reassess standard DIF analysis procedures using the MH statistic alone.

We particularly recommend that for easy and hard items STD-EISDIF data as well as MH data be taken into account during IA. Perhaps procedures such as differential distractor functioning (DDF) could be implemented, as DDF examines more than just the key, taking into account DIF on the distractors as well (Green, Crone, & Folk, 1989). Other methods of DIF analysis have also been investigated and proposed: Mapuranga, Dorans, and Middleton (2008) explored a variety of methods indexed by procedural complexity, computational intensity, and cost. That said, from a practical perspective, none of these proposed methods is a perfect fix. It seems clear that devising new and efficient methods of DIF analysis remains a compelling avenue for further research.

Recommendation 2: Carefully consider the amount of context that is provided with an item rather than just the particular content of an item without consideration of context.

This seems particularly relevant for material that has a history of differential item performance. As noted in the *ETS Guidelines for Fairness Review of Assessments* (ETS, 2009), “it may be appropriate to use a difficult word if ... its meaning is made clear by context.” By extension, it may be appropriate to use what might otherwise be considered a *moderately specialized* or *more than moderately specialized* passage to assess reading comprehension skills if enough context is provided for test takers to use logical reasoning to answer the related questions, regardless of the test taker's interest in, prior study of, or general familiarity with the topic.

The same cannot be said about certain kinds of SNCPs. Ranging in length from 15 to 30 words, SNCPs appear not to provide enough context for some test takers to sort through certain specialized language, or to use logic to deduce the correct response when the topic is unfamiliar. This seems particularly true when an item is testing the knowledge of words that have multiple meanings across different contexts. One observation of this study was that placing science vocabulary into a nonscience context in an SNCP sometimes resulted in differential responses among matched males and females (as compared to science vocabulary tested in a purely science context). By contrast, a *more than moderately specialized* science passage in this study provided enough context in its 110 words for most test takers to select the correct answer for each question, regardless of gender. The items accompanying the passage were not unduly hard, exhibited no DIF, and did not appear to generate affective sources of variance; that is, negative emotional responses appeared not to interfere with performance, as assessed through our cognitive interviews.



Indeed, it is clear from both the quantitative and qualitative phases of this study that gender differences in performance diminish when more, rather than less, context is provided for each item. A word or phrase that might come across as confusing or specialized in a 15- to 30-word SNCP is made less so when used in the context of a full reading passage—and items associated with the longest reading passages exhibit the smallest differences in performance among matched groups of males and females, regardless of perceived specialization of topic. This finding is in keeping with prior research on gender differences in performance, much of which was conducted on antonym and analogy item types, which used even fewer words and provided even less context than SNCPs (see Carlton & Harris, 1992; Curley & Schmitt, 1993; Schmitt & Bleistein, 1987).

**Recommendation 3:** When possible, use data to make decisions about test content believed to generate differential item functioning rather than presume bias based on gender.

Both the quantitative and qualitative results of this study indicate that sports content is no longer “inside baseball” that will advantage male test takers when matched for ability, despite widespread institutional belief to the contrary. Even with the limitation of the number of items in the quantitative phase and the number of participants in the qualitative phase, it was clear that sports content is, on the whole, appropriate for skills testing. Items with sports content no longer generate the significant DIF that was observed when similar studies were conducted in the 1980s and 1990s. Instead, both male and female participants in the qualitative study expressed pleasure and excitement when encountering sports material in the context of a test, and performance, when matched for ability, was mostly comparable. For those sports items that did generate significant DIF—whether measured by MH, STD-EISDIF, or both—other factors seemed to be in play, such as the potentially confounding effect of statistics in the short baseball passage on Wee Willie Keeler or the confluence of scientific and military terminology in the SNCPs about camouflage. It was outside the scope of this particular study to examine multidimensional DIF; it is an area that would benefit from further investigation into why such complexity results in differences that are not observed when the same content is tested on its own.

That is not to say that all sports and science passages are fair and valid when it comes to skills testing. Professional judgment and, when possible, statistical data gathered from the population that will take the examination remain critical in determining the appropriateness of items whose content has a history of differential performance for that population. We can have confidence, however, that sports context alone is no longer likely to elicit affective sources of construct-irrelevant variance. We also believe that providing increased context for answering both sports and science items will enhance the likelihood that the items will be fair and valid by providing a sound basis for reasoning and deduction, allowing test takers to understand, analyze, evaluate, and draw inferences from reading passages about varying topics, regardless of interest in or familiarity with the topic. In the end, questions about who's on first and what's on second are perhaps best answered by providing increased information. As in the famous Abbot and Costello comedy routine about baseball, one-word answers too easily mislead—and in a testing environment, that's no joke.

## Acknowledgments

The authors would like to thank Neil Dorans, Elizabeth Stone, Brent Bridgeman, Maria Elena Oliveri, Don Powers, and Michael Walker for their reviews of this report; Michael Zieky for providing historical information; Anthony Betancourt, Julia Farnum, Marcel Ionescu, Lauren Phelps, and Waverly Van Winkle for conducting the cognitive interviews; Elizabeth Ayad, Sean Bailey, Chelsea Ezzo, and Kurt Schmenger for coding the data; Kitty Sheehan and Rebecca Sundt for their assistance with APA style; Kathy Hassler for formatting the sample tests; and the College Board for facilitating data collection. Funding for this project was provided by ETS's Validity Initiative from 2009 to 2013.

## References

- Bleistein, C. A., Schmitt, A. P., & Curley, W. E. (1990). *Factors hypothesized to affect the performance of Black examinees on SAT-Verbal analogy items*. Paper presented at the annual meeting of the National Council on Measurement in Education, Boston, MA.
- Bleistein, C. A., & Wright, D. (1987). Assessment of unexpected differential difficulty for Asian-American candidates on the SAT. In A. P. Schmitt & N. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (Research Memorandum No. RM-87-01). Princeton, NJ: Educational Testing Service.



- Buck, G., Kostin, I., & Morgan, R. (2002). *Examining the relationship of content to gender-based performance differences in advanced placement exams* (Research Report No. RR-02-25). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2002.tb01892.x>
- Carlton, S. T., & Harris, A. M. (1992). *Characteristics associated with differential item functioning on the Scholastic Aptitude Test: Gender and majority/minority group comparisons* (Research Report No. RR-92-64). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1992.tb01495.x>
- Cole, N. (1997). *The ETS Gender Study: How females and males perform in educational settings*. Princeton, NJ: Educational Testing Service.
- Curley, W. E., & Schmitt, A. P. (1993). *Revising SAT-Verbal items to eliminate differential item functioning* (College Board Report No. 93-2). New York, NY: College Entrance Examination Board.
- Doolittle, A. E., & Welch, C. (1989). *Gender differences in performance on a college-level achievement test* (ACT Research Report Series No. 89-9). Iowa City, IA: American College Testing Program.
- Dorans, N. J. (1989). Two new approaches to assessing differential item functioning: Standardization and the Mantel-Haenszel method. *Applied Measurement in Education*, 2, 217-233.
- Dorans, N. J. (2013). *ETS contributions to the quantitative assessment of item, test, and score fairness* (Research Report No. RR-13-27, SPC-13-04). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2013.tb02334.x>
- Dorans, N. J., & Holland, P. W. (1992). *DIF detection and description: Mantel-Haenszel and standardization* (Research Report No. RR-92-10). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.1992.tb01440.x>
- Dorans, N. J., & Kulick, E. (1983). *Assessing unexpected differential item functioning performance of female candidates on SAT and TSWE forms administered in December 1977: An application of the standardization approach* (ETS Research Report No. RR-83-9). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2330-8516.1983.tb00009.x>
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355-368.
- Dorans, N. J., Schmitt, A. P., & Bleistein, C. A. (1992). The standardization approach to assessing comprehensive differential item functioning. *Journal of Educational Measurement*, 29, 309-319.
- Educational Testing Service. (2009). *ETS Guidelines for Fairness Review of Assessments*. Princeton, NJ: Author.
- Green, B. F., Crone, C. R., & Folk, V. G. (1989). A method for studying differential distractor functioning. *Journal of Educational Measurement*, 26, 147-160.
- Halpern, D. F. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Lawrence Erlbaum.
- Holland, P. W., & Thayer, D. T. (1988). *Differential item functioning and the Mantel-Haenszel procedure* (Research Report No. RR-86-89). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2330-8516.1986.tb00186.x>
- Lawrence, I. M., & Curley, W. E. (1989). *Differential item functioning for males and females on SAT-Verbal Reading subscore items: Follow-up study* (Research Report No. RR-89-22). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2330-8516.1989.tb00137.x>
- Lawrence, I. M., Curley, W. E., & McHale, F. J. (1988). *Differential item functioning for males and females on SAT-Reading subscore items* (Research Report No. RR-88-10). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2330-8516.1988.tb00266.x>
- Mapuranga, R., Dorans, N., & Middleton, K. (2008). *A review of recent developments in differential item functioning* (Research Report No. RR-08-43). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/j.2333-8504.2008.tb02129.x>
- Petersen, N. (1988). *DIF procedures for use in statistical analysis*. Unpublished manuscript.
- SAT Update. (2011, October 11). *SAT Critical Reading test development committee meeting* [PowerPoint slides]. New York, NY: College Board.
- Schmitt, A. P. (1988). Language and culture characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 25, 1-13.
- Schmitt, A. P., & Bleistein, C. A. (1987). *R* <http://dx.doi.org/10.1002/j.2330-8516.1987.tb00227.x>
- Schmitt, A. P., Curley, W. E., Bleistein, C. A., & Dorans, N. J. (1988). *Experimental evaluation of language and interest factors related to differential item functioning for Hispanic examinees on the SAT-Verbal*. New Orleans, LA: National Council on Measurement in Education.
- Schmitt, A. P., & Dorans, N. (1990). Differential item functioning for minority examinees on the SAT. *Journal of Educational Measurement*, 27, 67-81.
- Willingham, W., & Cole, N. (1997). *Gender and fair assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Women's Sports Foundation. (2008, August 20). *Statistics—Gender equity in high school and college athletics: Most recent participation and budget statistics*. Retrieved from <http://www.womenssportsfoundation.org/>
- Wright, D. (1987). An empirical comparison of the Mantel-Haenszel and standardization methods of detecting differential item performance. In A. P. Schmitt & N. Dorans (Eds.), *Differential item functioning on the Scholastic Aptitude Test* (Research Memorandum No. RM-87-01). Princeton, NJ: Educational Testing Service.

## Appendix A: Sample Test One



**TEST ONE**  
18 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ☒

- Dylan studied biology in college because he was fascinated with ----- life, creatures too small to be seen with the naked eye.  
(A) anatomical (B) omnivorous  
(C) microscopic (D) artificial  
(E) extraterrestrial
- Performance on the golf course is often -----: a professional golfer can play 18 holes in ten under par one day and fail to break par the next day.  
(A) enjoyable (B) profitable  
(C) inconsistent (D) uninteresting  
(E) deliberate
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular  
(C) informal (D) contrasting  
(E) occasional
- Up to twelve feet long and weighing more than 1,000 pounds, the leopard seal is surprisingly -----, moving quickly and gracefully.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the pressure that ice exerts on Earth's surface is -----, the rock beneath glaciers and ice sheets is pushed downward.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Mary Ann often hides her anger or distrust behind declarations of friendship; this ----- makes discerning what she really thinks a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Known for his ----- style of play, James did not hesitate to bodycheck opponents into the boards as he shot the puck across the ice toward the hockey goal.  
(A) methodical (B) inconsistent  
(C) cunning (D) adaptable  
(E) physical
- By the mid-1920s, the United States had more than 1,500 amusement parks; however, the stupendous Coney Island, dubbed "America's Playground," ----- all the others.  
(A) eclipsed (B) calibrated  
(C) systematized (D) reproduced  
(E) orbited

**GO ON TO THE NEXT PAGE** 



Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

Line In the early days of baseball, great hitters could succeed  
5 in ways they cannot now. "Wee Willie" Keeler could  
"hit 'em where they ain't" (and accrue an astonishing  
5 .432 batting average in 1897) because opposing players  
didn't yet know where they should stand. Consider the  
predicament of a great modern player. Baseball is now  
a science as much as a game. Every pitch is charted,  
every hit mapped to the nearest square inch. Defense  
has improved dramatically. A superb modern hitter like  
10 Wade Boggs was probably just as talented as Keeler, but  
the standard of play increased so much that Boggs could  
not take advantage of his opponents' weaknesses.

9. The passage serves primarily to

- (A) express nostalgia for a bygone time
- (B) condemn changes in the way a game is played
- (C) describe a significant transformation in a sport
- (D) advocate the continuation of a traditional practice
- (E) celebrate the achievements of two talented athletes

10. It can be inferred from the passage that the author believes that Keeler

- (A) succeeded because his opponents lacked the physical talents of contemporary athletes
- (B) was more inherently athletic than modern baseball players
- (C) used scientific insights to gain an advantage over other players
- (D) would not have had as good a batting average if he had played in Boggs's era
- (E) might have performed even better had he played today

Question 11 is based on the following passage.

Line Fairy tales, unlike other forms of literature, direct  
5 children to discover their identities, and also suggest  
what experiences are needed to develop their char-  
acters further. Fairy tales intimate that a rewarding,  
5 good life is within reach despite adversity—but only  
if one does not shy away from the hazardous struggles  
without which one cannot achieve true identity. These  
stories promise that if children engage in these fearsome  
and taxing searches, benevolent powers will come to the  
10 children's aid, and they will succeed. The stories also  
warn that those who are too timorous and narrow-minded  
to take these risks must settle for a humdrum existence—  
if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

GO ON TO THE NEXT PAGE 



Questions 12-18 are based on the following passage.

*This passage discusses the Heisenberg uncertainty principle, which holds that the position and momentum of any elementary particle cannot be precisely known at the same time. Heisenberg's formulation is a core component of quantum mechanics, the study of matter and energy at the atomic level.*

The uncertainty principle tells us that the universe is a frenetic place when examined on smaller and smaller distances and shorter and shorter time scales. If we try to pinpoint the location of elementary particles such as electrons\* by shining light of ever higher frequency on them, we measure their position with ever greater precision, but at a cost, since our observations become ever more disruptive. The high-frequency photons comprising the light have a lot of energy and therefore give the electrons a sharp "kick," significantly changing their velocities. Like the frenzy in a room full of children all of whose momentary positions you know with great accuracy but over whose velocities—the speeds and directions in which they are moving—you have almost no control, this inability to know both the positions and velocities of elementary particles implies that the microscopic realm is intrinsically turbulent.

Although this example conveys the basic relationship between uncertainty and frenzy, it actually reveals only part of the story. It might lead you to think, for instance, that uncertainty arises only when we clumsy observers of nature stumble onto the scene. This is *not* true. Even in the most quiescent setting imaginable, such as an empty region of space, the uncertainty principle tells us that from a microscopic vantage point there is a tremendous amount of activity. And this activity gets increasingly agitated on ever smaller distance and time scales.

Quantum accounting is essential to understand this. Just as you might temporarily borrow money to overcome an important financial obstacle, a particle such as an electron can temporarily borrow energy to overcome a literal physical barrier. But quantum mechanics forces us to take the analogy one important step further. Imagine a man who is a compulsive borrower and goes from friend to friend asking for money. The shorter the time for which a friend can lend him money, the larger the loan he seeks. Borrow and return, borrow and return—over and over again with unflagging intensity he takes in money only to give it back in short order. The amount of money the compulsive borrower possesses at any given moment goes through extreme fluctuations, but when all is said and done, an accounting of his finances shows that he is no better off than when he began.

Heisenberg's uncertainty principle asserts that a similar frantic shifting back and forth of energy and momentum is occurring perpetually in the universe on microscopic distance and time intervals. Even in an empty region of space—inside an empty box, for example—the uncertainty principle says that the energy and momentum are *uncertain*: They fluctuate between extremes that get larger as the size of the box and the time scale over which it is examined get smaller and smaller. It's as if the region of space inside the box is a compulsive "borrower" of energy and momentum, constantly extracting "loans" from the universe and subsequently "paying" them back. But what participates in these exchanges in, for instance, a quiet, *empty* region of space? Everything. Literally. Energy (and momentum as well) is the ultimate convertible currency.  $E=mc^2$  tells us that energy can be turned into matter and vice versa. Thus if an energy fluctuation is big enough, it can momentarily cause, for instance, an electron and its antimatter companion the positron to erupt into existence, even if the region had been initially empty! Since this energy must be quickly repaid, these particles will annihilate one another after an instant, relinquishing the energy borrowed in their creation. And the same is true for all of the other forms that energy and momentum can take—other particle eruptions and annihilations, wild electromagnetic-field oscillations, weak and strong force-field oscillations. Quantum-mechanical uncertainty tells us the universe is a teeming, chaotic, frenzied arena on microscopic scales.

*\*Elementary particles, such as electrons and photons, are particles so small that they cannot be seen by the naked eye.*

12. The first sentence (lines 1-3) implies that a large-scale examination of the universe would decrease the appearance of
  - (A) comprehensiveness
  - (B) uniformity
  - (C) frenzy
  - (D) opacity
  - (E) stability
13. The author would most likely characterize information resulting from the "observations" (line 7) as
  - (A) accidentally accurate
  - (B) regrettably useless
  - (C) unnecessarily exhaustive
  - (D) fancifully ephemeral
  - (E) necessarily limited

GO ON TO THE NEXT PAGE





14. Lines 8-10 (“The high-frequency . . . velocities”) indicate that one way in which photons can affect electrons is by
- (A) increasing their mass
  - (B) altering their speed
  - (C) reducing their light
  - (D) obscuring their position
  - (E) decreasing their numbers
15. The author would most likely agree that although “we clumsy observers” (line 21) can
- (A) theorize microscopic uncertainty, we are unable to prove it
  - (B) negate microscopic uncertainty, we are unable to create it
  - (C) predict microscopic uncertainty, we are unable to understand it
  - (D) forestall microscopic uncertainty, we are not fundamentally its target
  - (E) recognize microscopic uncertainty, we are not necessarily its cause
16. In terms of overall energy and momentum, a “region of space” (lines 52-53) subject to the “accounting” mentioned in line 42 would appear
- (A) massive, despite its small size
  - (B) depleted, despite its turbulent state
  - (C) chaotic, despite the phenomenon of uncertainty
  - (D) static, despite its intense atomic-level activity
  - (E) undisturbed, despite its continual growth
17. In lines 55-63 (“But what . . . empty”), the author emphasizes that energy can be
- (A) expressed as forces
  - (B) produced by motion
  - (C) used for heat
  - (D) transferred by currents
  - (E) transformed into matter
18. An important implication of the passage as a whole is that
- (A) scientific theories are unsupportable
  - (B) perceptions may not always be accurate
  - (C) experiments should be carefully designed
  - (D) technology leads to theoretical understanding
  - (E) much of the universe will remain unexamined

**S T O P**

## Appendix B: Sample Test Two

2    □    □    2    □    □    2

## TEST TWO

18 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

1. The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
2. With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
3. Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
4. Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
5. Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
6. Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
7. Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
8. Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

GO ON TO THE NEXT PAGE

# 2      □      □      2      □      □      2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

**Questions 9-10 are based on the following passage.**

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

**9.** The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

**10.** In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

**Question 11 is based on the following passage.**

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

**11.** The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

**GO ON TO THE NEXT PAGE** 



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the two of us together because neither of us likes to back down.

20 "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it, pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

\*national competition for amateur boxing

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight' "), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill



17. In line 58, the repetition of the word “*chuddica*” captures the

- (A) pattern of the author’s footwork
- (B) sound that the punching bag makes
- (C) rhythm of the author’s labored breathing
- (D) cycle of the author’s thoughts
- (E) ticking of the time clock

18. The role of the bell in the last paragraph is primarily to

- (A) imitate an audience’s reaction to a fight
- (B) give the author’s opponent a needed break
- (C) prompt the author to stretch and drink water
- (D) denote changes in the author’s mood
- (E) mark the progress of the author’s training session

**S T O P**

**Appendix C: Protocol for Interviews****ID Information**

Student ID: \_\_\_\_\_

Form:   2  

Test Adm Initials: \_\_\_\_\_

Date: \_\_\_\_\_

School: \_\_\_\_\_

# SAT Critical Reading COGNITIVE INTERVIEW PROTOCOL

## COGNITIVE LAB INTERVIEW PROTOCOL

### SEATING INSTRUCTIONS



Follow the school's procedures to escort one student from his or her classroom to an empty testing room or corner of a quiet room.

Make sure that you and the student are seated so that you face one another directly across the table or desk if possible. Set the interview protocol and the stopwatch in front of you and the closed student test booklet in front of the student. Place the digital recorder on the table near the student.

### INTRODUCTION

Hello! My name is (YOUR NAME). I work at Educational Testing Service. Today you will be participating in a research study on the SAT reading test. We are interested to learn your thoughts and opinions about the test.

Do you have any questions before I begin the instructions?



Create small talk...ask a couple friendly questions such as the following:

Do you like reading?



If yes, say

Oh, good, then I think you will like what we are doing today.



If no, say

Okay, well what is your favorite subject?



Appropriately respond. (Also make a note in the test administrator answer sheet whether student likes to read.)

Then move go back to the directions. Say,

You will not be asked to write your name on any of the work you do, and *no one* at your school will see your answers. Participating in this study is voluntary. If at any point you decide you do not want to continue that is your choice, and you are free to stop.

Because the study is so important, I want to thank you ahead of time for all the hard work you are about to put into your answers. Some of the questions might be difficult, but just do the best that you can.

#### AUDIO RECORDING/SOUND CHECK

We are going to record today's test so I remember what you say. First we're going to test out this audio recorder.



Make a test recording by telling the student:

When I say GO, I want you to say, "Today is Friday" in a clear voice.

Press the RECORD button on the audio recorder and say "YOU MAY BEGIN" to prompt the student.

Be sure to press RECORD *before* you tell the student to begin so your voice is recorded too.

If necessary, remind the student to say the sentence ("Today is Friday.") Once the student has said the sentence, press PLAY to listen to the recording.

- If you *cannot* hear the student clearly on the recorder repeat the test recording process until you get a clear recording, then continue with the administration:
- If you *can* hear the student clearly, continue with the administration:

Good job. Now we'll begin the instructions.



**DIRECTIONS****Think aloud hints to administrator**

We're interested in capturing all the mental processes that the student engages while completing these math items. To do this your goal is to have the student speak aloud all his or her thoughts while solving the items and by asking follow up questions for each item. There are several things to keep in mind to ensure the data collected is as complete as possible:

- If a student is continually providing short responses or not answering, use "continuers" to encourage the student to be more descriptive. The trick is to get them to continue without putting words in their mouth. Make sure not to ask questions that lead a student's response. You have to be as objective and non-bias as possible. The way you get them to keep talking is by offering a gentle verbal nudge like this:

"What are you thinking now?"

"Any other thoughts?"

"Tell me how you came to pick that answer."

- Use your best judgment. If a student is responsive but is not easily explaining his or her reasoning, gently probe the student without putting any bias in their response.
- Focus on the task at hand, the particular item. Do not try to ask a student a question in general terms.



**INSTRUCTIONS****DIRECTIONS**

Please open your test booklet to the first page where you will see the directions. To mark an answer, circle the letter next to your choice.



Direct student to the first page and point to the directions.

**PRACTICE EXAMPLE AND DETAILS**

Today you will respond to some SAT sentence completion questions and reading passages with multiple choice questions. To better understand what you are thinking, I want you to think aloud all your thoughts as you read the questions and passages, and respond to the questions. In a moment, I will show you how to think out loud as if I were a student,

Please turn the page in your booklet so you are at the first practice page.



Make sure student is on the administrator practice page.

You can follow along in your book to see the question I will read aloud and will think aloud as I come to my answer. When I am finished, you will get a chance to practice.



Make sure student is on the correct page.

Read item and answer choices aloud and demonstrate think aloud

## Interviewer practice think aloud

## Directions

Read the sentence. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

Although some think the terms "bug" and "insect" are -----, the former term actually refers to ----- group of insects.

- ☐ (A) parallel . . an identical
- ☐ (B) precise . . an exact
- ☐ (C) interchangeable . . a particular
- ☐ (D) exclusive . . a separate
- ☐ (E) useful . . a useless

- As you heard, I was talking the whole time I was working on the question.
- I began by reading the question and the answer choices aloud, and then
- I said aloud everything I was thinking about as I answered the question.
- I also made a mistake, which is okay because sometimes the answer doesn't come quickly to me and I have to try several things before I can answer the question.

Do you have any questions?



Answer any questions the student may ask

Now I'd like you to practice answering a question by thinking aloud. Please read the question and the answer choices aloud first, and then I'd like you to say all of your thoughts aloud as you answer the question.



Direct student to the student practice question

## Student practice think aloud

The addition of descriptive details to the basic information serves to ----- the book by producing a fuller account.

- ☐ (A) invalidate
- ☐ (B) objectify
- ☐ (C) incite
- ☐ (D) celebrate
- ☐ (E) enrich

[According to the student's performance, assist with appropriate suggestions for more accurate thinking aloud or say] Good job! Do you have any questions?



Answer any questions the student may ask.

Okay, let's move on to the actual test questions. I want you to read each question and say all your thoughts out loud as you answer the question. Complete all the questions in the set or until I tell you to stop, then I will ask you some follow up questions. There are a total of 20 questions in this test. After you finish the 20 questions, I will ask you some final questions about this experience. So that I can remember what you say, I am going to be taking notes while you think aloud and answer the questions. This whole session will take 90 minutes.

**REMINDER:** Press the RECORD button to begin recording and continue with the administration.

[SAY] The student ID and the date.

Please turn the page in your booklet to the first test question and you may begin reading aloud and answering the question.

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 1

Item 2

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 1

Item 2

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the item topics.

Item 1

Item 2

**Other noteworthy comments:**

Item 1

Item 2



# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful
- (B) end . . divisive
- (C) overcome . . unattractive
- (D) extend . . satisfactory
- (E) resolve . . acceptable

(A) (B) (C) (D) ●

1. The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.
  - (A) revolve (B) alight (C) roam
  - (D) hover (E) hibernate
2. With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.
  - (A) dexterity (B) teamwork (C) opportunity
  - (D) stamina (E) independence
3. Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.
  - (A) confusing (B) unpopular (C) informal
  - (D) contrasting (E) occasional
4. Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.
  - (A) aggressive (B) leisurely (C) agile
  - (D) noisy (E) bulky
5. Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.
  - (A) undetectable (B) debatable
  - (C) considerable (D) imprecise
  - (E) variable
6. Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.
  - (A) concealment . . facile
  - (B) masquerade . . paltry
  - (C) camouflage . . formidable
  - (D) simulation . . fortuitous
  - (E) tractability . . confounding
7. Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.
  - (A) nondescript (B) mercurial (C) tractable
  - (D) puerile (E) rife
8. Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.
  - (A) automatic (B) periodic (C) seismic
  - (D) amorphous (E) elliptical

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 3

Item 4

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 3

Item 4

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the item topics.

Item 3

Item 4

**Other noteworthy comments:**

Item 3

Item 4



# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 5

Item 6

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 5

Item 6

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the item topics.

Item 5

Item 6

**Other noteworthy comments:**

Item 5

Item 6

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ☒

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 7

Item 8

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 7

Item 8

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the item topics.

Item 7

Item 8

**Other noteworthy comments:**

Item 7

Item 8

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Now I'm going to ask you some questions about the questions you just answered.

Think back to **question 1**

#### Retrospective Questions

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## TEST TWO 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful
- (B) end . . divisive
- (C) overcome . . unattractive
- (D) extend . . satisfactory
- (E) resolve . . acceptable

(A) (B) (C) (D) ●

1. The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.
  - (A) revolve (B) alight (C) roam
  - (D) hover (E) hibernate
2. With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.
  - (A) dexterity (B) teamwork (C) opportunity
  - (D) stamina (E) independence
3. Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.
  - (A) confusing (B) unpopular (C) informal
  - (D) contrasting (E) occasional
4. Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.
  - (A) aggressive (B) leisurely (C) agile
  - (D) noisy (E) bulky
5. Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.
  - (A) undetectable (B) debatable
  - (C) considerable (D) imprecise
  - (E) variable
6. Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.
  - (A) concealment . . facile
  - (B) masquerade . . paltry
  - (C) camouflage . . formidable
  - (D) simulation . . fortuitous
  - (E) tractability . . confounding
7. Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.
  - (A) nondescript (B) mercurial (C) tractable
  - (D) puerile (E) rife
8. Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.
  - (A) automatic (B) periodic (C) seismic
  - (D) amorphous (E) elliptical

Think back to **question 2**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Think back to **question 3**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful
- (B) end . . divisive
- (C) overcome . . unattractive
- (D) extend . . satisfactory
- (E) resolve . . acceptable

(A) (B) (C) (D) ●

1. The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.
  - (A) revolve (B) alight (C) roam
  - (D) hover (E) hibernate
2. With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.
  - (A) dexterity (B) teamwork (C) opportunity
  - (D) stamina (E) independence
3. Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.
  - (A) confusing (B) unpopular (C) informal
  - (D) contrasting (E) occasional
4. Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.
  - (A) aggressive (B) leisurely (C) agile
  - (D) noisy (E) bulky
5. Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.
  - (A) undetectable (B) debatable
  - (C) considerable (D) imprecise
  - (E) variable
6. Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.
  - (A) concealment . . facile
  - (B) masquerade . . paltry
  - (C) camouflage . . formidable
  - (D) simulation . . fortuitous
  - (E) tractability . . confounding
7. Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.
  - (A) nondescript (B) mercurial (C) tractable
  - (D) puerile (E) rife
8. Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.
  - (A) automatic (B) periodic (C) seismic
  - (D) amorphous (E) elliptical

Think back to **question 4**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Think back to **question 5**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Think back to **question 6**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Think back to **question 7**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## TEST TWO

### 20 Questions

**Directions:** For each question in this section, select the best answer from among the choices given and fill in the corresponding circle on the answer sheet.

Each sentence below has one or two blanks, each blank indicating that something has been omitted. Beneath the sentence are five words or sets of words labeled A through E. Choose the word or set of words that, when inserted in the sentence, best fits the meaning of the sentence as a whole.

**Example:**

Hoping to ----- the dispute, negotiators proposed a compromise that they felt would be ----- to both labor and management.

- (A) enforce . . useful  
(B) end . . divisive  
(C) overcome . . unattractive  
(D) extend . . satisfactory  
(E) resolve . . acceptable

(A) (B) (C) (D) ●

- The rapid beat and motion of its wings allow a hummingbird to ----- while it feeds, so that it appears suspended in midair.  
(A) revolve (B) alight (C) roam  
(D) hover (E) hibernate
- With an official distance of more than 42 kilometers, the marathon is a true test of -----: it takes most runners many grueling hours to complete.  
(A) dexterity (B) teamwork (C) opportunity  
(D) stamina (E) independence
- Among the first Chinese American writers, sisters Edith and Winnifred Eaton had ----- literary approaches: Edith wrote social novels set in the United States, while Winnifred wrote romantic stories set in Japan.  
(A) confusing (B) unpopular (C) informal  
(D) contrasting (E) occasional
- Fluent in several languages and knowledgeable about many subjects, Pat has an amazingly ----- mind, perceptive and quick.  
(A) aggressive (B) leisurely (C) agile  
(D) noisy (E) bulky
- Because the popularity of jazz music over the past century has been -----, jazz's influence on other art forms is now apparent.  
(A) undetectable (B) debatable  
(C) considerable (D) imprecise  
(E) variable
- Frogfish have the ability to change color and blend into the background of their habitat; this ----- often makes locating them a ----- task.  
(A) concealment . . facile  
(B) masquerade . . paltry  
(C) camouflage . . formidable  
(D) simulation . . fortuitous  
(E) tractability . . confounding
- Injuries are ----- among short-track skaters: they occur so routinely that short track is often described as roller derby on ice.  
(A) nondescript (B) mercurial (C) tractable  
(D) puerile (E) rife
- Political pundits described the shift in voting patterns as nothing less than -----, so strong and widespread was its impact.  
(A) automatic (B) periodic (C) seismic  
(D) amorphous (E) elliptical

Think back to **question 8**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

Great, thank you, now I'd like you to move to the next section. Please read aloud the passage on the left and think aloud as you answer the 2 questions about it.

# 2      □      □      2      □      □      2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 9

Item 10

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 9

Item 10

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 9

Item 10

**Other noteworthy comments:**

Item 9

Item 10

# 2      □      □      2      □      □      2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

Now I'm going to ask you some questions about the questions you just answered.

Think back to **question 9**

### **Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2      □      □      2      □      □      2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

Think back to **question 10**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

Now I'd like you to think about the passage you just read about Cnidaria.

1. Do you know much about this topic?
2. How much did you like reading about this topic?
3. Have you read anything else on this topic before?
  - a. If yes, for school/pleasure or both?
4. Is there anything about the passage that confused you?
5. Did answering the questions help you understand the passage better? Please explain.
6. Is there anything else you want to say about your interest in this passage or this topic in general?

Great, thank you, now I'd like you to move to the next passage. Please read aloud the passage on the right and think aloud as you answer the 1 question about it.

# 2      □      □      2      □      □      2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 11

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 11

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 11

**Other noteworthy comments:**

Item11



Interviewer reminder – no retrospective questions asked for item 11.



# 2 □ □ 2 □ □ 2

Each passage below is followed by questions based on its content. Answer the questions on the basis of what is stated or implied in each passage and in any introductory material that may be provided.

Questions 9-10 are based on the following passage.

The phylum Cnidaria, with a silent “c,” includes jellyfish, sea anemones, and corals. A key feature of cnidarians is the presence of specialized stinging cells called nematocysts. Employed in feeding and defense, nematocysts act like miniature harpoons, shooting out a dart tethered by a thread that can harpoon prey or enemies. These rather sophisticated structures belie the simple organization of the cnidarian body. Cnidarians lack the coeloms (fluid-filled body cavities) found in most invertebrates. Instead, their tissue organization is diploblastic: the walls of the body possess two layers. Between these walls is a mesoglea comprising stringy cells and a nerve net. Viewed from above, cnidarians usually exhibit radial symmetry.

9. The primary purpose of the passage is to

- (A) compare the behavior of jellyfish, sea anemones, and corals
- (B) question the adaptive features of jellyfish, sea anemones, and corals
- (C) examine the role of nematocysts in the phylum Cnidaria
- (D) highlight some important distinctions among members of the phylum Cnidaria
- (E) describe particular characteristics of members of the phylum Cnidaria

10. In the passage, the “organization” (line 10) is described primarily to show how it

- (A) is used by cnidarians to control the flow of fluids
- (B) is consistent with cnidarians’ methods of feeding and defense
- (C) reveals the need for cnidarians to have multiple layers of tissue
- (D) distinguishes cnidarians from most other invertebrates
- (E) helps cnidarians navigate through their environment

Question 11 is based on the following passage.

Fairy tales, unlike other forms of literature, direct children to discover their identities, and also suggest what experiences are needed to develop their characters further. Fairy tales intimate that a rewarding, good life is within reach despite adversity—but only if one does not shy away from the hazardous struggles without which one cannot achieve true identity. These stories promise that if children engage in these fearsome and taxing searches, benevolent powers will come to the children’s aid, and they will succeed. The stories also warn that those who are too timorous and narrow-minded to take these risks must settle for a humdrum existence—if an even worse fate does not befall them.

11. The primary purpose of the passage is to

- (A) argue that children find fairy tales more appealing than contemporary stories
- (B) demonstrate how fairy tales serve different needs in different cultures
- (C) show how fairy tales help prepare children for the challenges of the real world
- (D) caution that fairy tales may have a negative effect on children
- (E) indicate that fairy tales are an entertaining way to escape reality

Now I'd like you to think about the passage you just read about fairy tales.

7. Do you know much about this topic?
8. How much did you like reading about this topic?
9. Have you read anything else on this topic before?
  - a. If yes, for school/pleasure or both?
10. Is there anything about the passage that confused you?
11. Did answering the questions help you understand the passage better? Please explain.
12. Is there anything else you want to say about your interest in this passage or this topic in general?

Great, thank you, you may move to the next passage. Please read aloud the passage and think aloud as you answer the questions about it.

# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

\* national competition for amateur boxing

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight' "), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 12

Item 13

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 12

Item 13

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 12

Item 13

**Other noteworthy comments:**

Item 12

Item 13



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

\* national competition for amateur boxing

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight' "), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 14

Item 15

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 14

Item 15

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 14

Item 15

**Other noteworthy comments:**

Item 14

Item 15



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

\* national competition for amateur boxing

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight' "), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 16

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 16

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 16

**Other noteworthy comments:**

Item 16



17. In line 58, the repetition of the word “*chuddica*” captures the

- (A) pattern of the author’s footwork
- (B) sound that the punching bag makes
- (C) rhythm of the author’s labored breathing
- (D) cycle of the author’s thoughts
- (E) ticking of the time clock

18. The role of the bell in the last paragraph is primarily to

- (A) imitate an audience’s reaction to a fight
- (B) give the author’s opponent a needed break
- (C) prompt the author to stretch and drink water
- (D) denote changes in the author’s mood
- (E) mark the progress of the author’s training session

**Concurrent reporting notes page**

**ITEM CONTENT and ANSWERING STRATEGY:** Take notes on student's ability to answer correctly. Include any information on any clear indication of understanding, any noticeable errors in the concept, misunderstanding of what the question is asking, or misunderstanding with some or all of the answer choices, or specific vocabulary issues. Also, takes notes on whether the student is in guessing mode, problem solving mode or somewhere in between. Lastly, summarize how student solved the problem, the steps that he or she is taking.

Item 17

Item 18

**ITEM CORRECTNESS** Take notes on any observed confidence or lack of confidence in the student's selection of the correct answer choice. Does the student think his or her answer is correct? How certain is he or she that his or her choice is correct?

Item 17

Item 18

**ITEM INTEREST** Take notes on any observed interest or lack of interest in the passage or item topics. Identify whether the interest/lack of interest is in the passage or specific items.

Item 17

Item 18

**Other noteworthy comments:**

Item 17

Item 18

# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

*\*national competition for amateur boxing*

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill



Now I'm going to ask you some questions about the questions you just answered.

Think back to **question 12**

### Retrospective Questions

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

*\*national competition for amateur boxing*

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

Think back to **question 13**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)

# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

*\*national competition for amateur boxing*

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot!*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

Think back to **question 14**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

*\*national competition for amateur boxing*

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

Think back to **question 15**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

## Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

*\*national competition for amateur boxing*

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

Think back to **question 16**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



17. In line 58, the repetition of the word “*chuddica*” captures the

- (A) pattern of the author’s footwork
- (B) sound that the punching bag makes
- (C) rhythm of the author’s labored breathing
- (D) cycle of the author’s thoughts
- (E) ticking of the time clock

18. The role of the bell in the last paragraph is primarily to

- (A) imitate an audience’s reaction to a fight
- (B) give the author’s opponent a needed break
- (C) prompt the author to stretch and drink water
- (D) denote changes in the author’s mood
- (E) mark the progress of the author’s training session

Think back to **question 17**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



17. In line 58, the repetition of the word “*chuddica*” captures the

- (A) pattern of the author’s footwork
- (B) sound that the punching bag makes
- (C) rhythm of the author’s labored breathing
- (D) cycle of the author’s thoughts
- (E) ticking of the time clock

18. The role of the bell in the last paragraph is primarily to

- (A) imitate an audience’s reaction to a fight
- (B) give the author’s opponent a needed break
- (C) prompt the author to stretch and drink water
- (D) denote changes in the author’s mood
- (E) mark the progress of the author’s training session

Think back to **question 18**

**Retrospective Questions**

1. Was this question hard, in the middle or easy for you?
  - a. Why?
  
2. Is this something you have learned in class before?  
if Yes, when did you learn it?
  
3. Were there any words, ideas, or anything else that made it easy to answer this question?  
Yes or No (*indicate the words if Yes*)
  
4. Were there any words, ideas, or anything else that made this question harder or confusing to answer this question? Yes or No (*indicate the words if Yes*)



# 2 □ □ 2 □ □ 2

Questions 12-18 are based on the following passage.

*This passage is adapted from a 2005 book about women's boxing. The author trained with the women about whom she wrote.*

How odd that I am the only one of us still boxing! I suppose it means I'm still listening for what it has to tell me. Its promise about what a woman can be is not yet something I can retain without the physical reminder.

Line 5 When I have to miss a few weeks at the gym, I feel an aspect of myself begin to fade, like a mirage, like something that was never real to begin with. I lose a few pounds; even my voice seems less substantial. I sleep fitfully, and in my dreams I waft like smoke. Each time

10 I return to the gym, it's as though I recover substance, ferocity, vigor.

Kimber and I keep sparring when we can. The last time I gave her another bloody nose and she gave me a black eye, and we stood side by side in the bathroom afterward,

15 washing the Vaseline and sweat off our faces. She offered me some of her apricot facial scrub. She wants to train hard this fall, to get ready for the Golden Gloves.\* I asked our coach, John, if he'd let me work her corner with him, and he said yes. He said nobody at the gym spars as hard as the

20 two of us together because neither of us likes to back down. "You know how they say, 'Styles make fights'? That's you and Kimber." He shakes his head, grinning. "You battle. You both like to stand in the pocket and fight."

Today, after saying goodbye to Raphi and her baby,

25 I go to the gym. It's a fine day, rare, a tang of autumn in the air. I have no appointment for a lesson; the gym is all but empty. I wrap my hands, put my water bottle on the windowsill, turn on the time clock, do three rounds of jump rope. The nylon rope smacks the floor like a firecracker.

30 I stretch, shadowbox three rounds in front of the mirror, my breath singing finely through my lips on every punch. I make a little story out of each round, not a story but a game plan, an imagined opponent and the footwork and weaving and punches I would employ. *Work tight,*

35 *work small.* John's voice in my head. *Throw it out straight, don't drop the arm. Turn the foot.* The time clock keeps sounding, pulling me out of the story, the physical narrative, setting me up for the next one. A drink of water, warm and sweet. The front of my shirt darkly speckled with

40 sweat. Walk around, walk around. My hands like glowing embers in their gloves. Wait for the bell. Bell. I work the double-end bag, moving in, moving out, circling it, circling back. Bob and weave. *Keep your chin down. Double jab. Stick it!* Stretching out between rounds. The spun gold

45 streaming in the open windows, the promise of coming apples, coming wood smoke. I strike the heavy bag. Two jabs and a right, push off out of range and then fast back in for another one, two, step in, three. Bell. Two more rounds on the heavy bag. I set it swinging, court and dodge it,

50 pound and pummel it, face it and pivot away. I work on the double hook John showed me: one to the body, one to the head. One, two, hook low, hook high. *Get in close! Commit!* Bell. My arms ache. I pull off my gloves, toss them toward my gym bag. I swig more water. I am wet all

55 over and bathed in heat, in saffron and yellow and rose.

*Finish with the speed bag.* John's routine. I go over to the tear-shaped bag, wait for the bell, work up a rhythm. *Chuddica chuddica chuddica.* The bag and my fists go to and fro, they meet and part, meet and part. The bag is a blur

60 I cannot see. It's as though all of myself has converged in my body. I might almost trust it to be true.

\*national competition for amateur boxing

12. The "physical reminder" (line 4) refers to the author's
  - (A) sensory experience of boxing
  - (B) nostalgic memories of boxing
  - (C) old injuries from boxing
  - (D) boxing trophies and souvenirs
  - (E) boxing garb and gear
13. In lines 19-23 ("He said . . . fight'"), John suggests that the two boxers are both
  - (A) stylish
  - (B) capricious
  - (C) antagonistic
  - (D) tenacious
  - (E) petulant
14. The "story" mentioned twice in line 32 refers most directly to the author's
  - (A) manuscript about the sport of boxing
  - (B) fantasy about being a professional boxer
  - (C) pretense that she is skilled at boxing
  - (D) match with an invented adversary
  - (E) internal monologue to help reinforce her self-esteem
15. The words in italics in lines 34-36 ("*Work . . . foot*") are best described as
  - (A) warnings
  - (B) suggestions
  - (C) directives
  - (D) commentary
  - (E) heckling
16. Lines 50-53 ("I . . . Bell") indicate that the author is
  - (A) considering a modified strategy
  - (B) inventing a new drill
  - (C) practicing a particular technique
  - (D) comparing a series of maneuvers
  - (E) displaying a universal skill

Now I'd like you to think about the passage you just read about Women's Boxing.

1. Do you know much about this topic?
2. How much did you like reading about this topic?
3. Have you read anything else on this topic before?
  - a. If yes, for school/pleasure or both?
4. Is there anything about the passage that confused you?
5. Did answering the questions help you understand the passage better? Please explain.
6. Is there anything else you want to say about your interest in this passage or this topic in general?

You are finished with the test. I have just a few more questions to ask you about your hobbies, interests and classes you have taken.

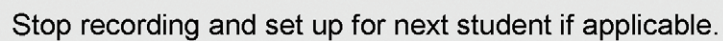
### Post Test Questions

- 1. Do you enjoy reading about science? Please explain.
  
  
  
  
  
  
  
  
  
  
- 2. Please name the science classes you've taken in high school and the grades you received in each class.

Science Course Name	Science Course Grade
1	
2	
3	
4	
5	
6	
7	
8	

- 3. Please describe any science-related interests and hobbies you have.
  - a. Explain how you engage in the science- related interests and hobbies you enjoy [e.g. read information online, in books, in magazines, member of one or more regional or school club, attend science-related events (explain) visit science museums, play science video games, attend any science-related camps, or any other way]

- Thank you very much! That is the end of the session today. I really appreciate all your hard work. As a thank you, here is a gift certificate for your time.



**Suggested citation:**

Chubbuck, K., Curley, W. E., & King, T. C. (2016). *Who's on first? Gender differences in performance on SAT® Critical Reading items with sports and science content* (ETS Research Report No. RR-16-26). Princeton, NJ: Educational Testing Service. <http://dx.doi.org/10.1002/ets2.12109>

**Action Editor:** Donald Powers

**Reviewers:** Maria Oliveri and Brent Bridgeman

ETS, the ETS logo, and MEASURING THE POWER OF LEARNING are registered trademarks of Educational Testing Service (ETS). SAT and PSAT/NMSQT are registered trademarks of the College Board. All other trademarks are property of their respective owners

Find other ETS-published reports by searching the ETS RESEARCHER database at <http://search.ets.org/researcher/>